REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL/01 AUG 91 TO 31 JUL 94

4. TITLE AND SUBTITLE

ADVANCED METHODS OF APPROXIMATE REASONING

5. FUNDING NUMBERS

2304/GS
F49620-91-C-0060

6. AUTHOR(S)

DR ENRIQUE RUSPINI

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SRI
ARTIFICIAL INTELLIGENCE CENTER
COMPUTING AND ENGINEERING SCEINCES DIVISION
333 RAVENSWOOD AVE
MENLO PARK, CA 94025-3493

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 95-0210

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NM
110 DUNCAN AVE, SUTE B115
BOLLING AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

F49620-91-C-0060

11. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

13. ABSTRACT (Maximum 200 words)

The major goals of the program of research on approximate reasoning of the Artificial Intelligence Center (AIC), SRI International, are:

The development of sound formal foundations to explain the different methodologies proposed to solve the problems associated with the processing of imprecise and uncertain information

The identification of criteria to determine the applicability of these methodologies to specific problems

The development of methods for the approximate modeling of real world systems

The development of techniques for the analysis of approximate models and determination of their behavioral properties, such as stability, robustness, and controllability.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR(SAME AS REPORT)

SRI International

Final Report • December 1994

Advanced Methods of Approximate Reasoning

SRI Project No. 2656
Contract No. F49620-91-C-0060

Covering the period 1 August 1992 to 30 September 1994

Prepared by:

Dr. Enrique H. Ruspini, Principal Investigator
Artificial Intelligence Center
Computing and Engineering Sciences Division

Prepared for:

United States Air Force Office of Scientific Research
Building 410
Bolling Air Force Base, D.C. 20332-6448
Attention: Dr. Abraham M. Waksman, Program Manager
Mathematical and Computer Sciences Directorate

(AFSC)
ed and is
0-12

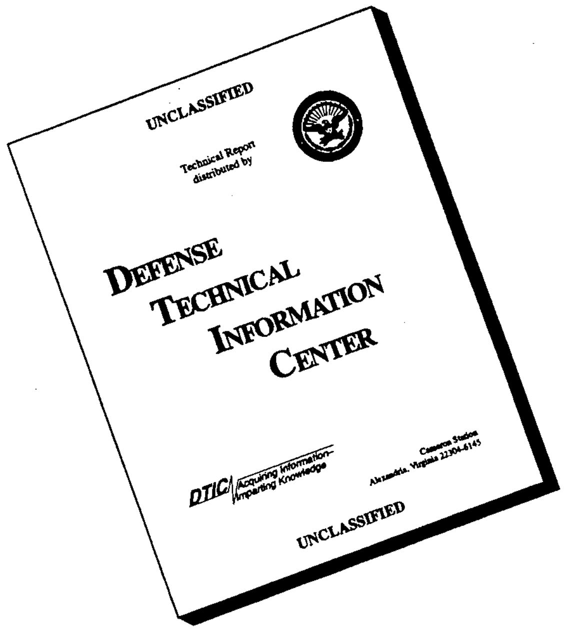
Approved by:

Dr. C. Raymond Perrault, Director
Artificial Intelligence Center

Dr. Donald L. Nielson, Vice President
Computing and Engineering Sciences Division

333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • (415) 326-6200 • FAX: (415) 326-5512 • Telex: 334486

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

PART I

1 Introduction

The major goals of the program of research on approximate reasoning of the Artificial Intelligence Center (AIC), SRI International, are

- The development of sound formal foundations to explain the different methodologies proposed to solve the problems associated with the processing of imprecise and uncertain information

- The identification of criteria to determine the applicability of these methodologies to specific problems

- The development of methods for the approximate modeling of real-world systems

- The development of techniques for the analysis of approximate models and determination of their behavioral properties, such as stability, robustness, and controllability

A major difference between our research goals and those of similar approaches to the treatment of imprecision and uncertainty is the development of techniques based on a sound understanding of the nature of *approximate models* of a real-world system. An approximate model is a qualitative representation of an aspect of reality developed to analyze and predict the behavior of a physical system. In the context of this discussion, the term "qualitative" is intended to indicate that the model may be based on vague, incomplete, imprecise, or uncertain information.

The need to rely on approximate models may be the consequence of poor understanding of system laws, imprecision and errors in measurements or observations, or practical limitations on our analytical capabilities. These limitations may require the consideration of simpler models that, avoiding unnecessary detail, attempt to focus only on significant aspects of system behavior.

We seek to develop sound analytical techniques to determine, on the basis of those approximate models, control and planning policies or important system properties such as "stability" or "robustness." We also seek useful characterizations of those properties in a qualitative context at a level of description similar to that used in the models themselves.

Specific Research Objectives of This Program

Our investigative efforts under the sponsorship of the U.S. Air Force Office of Scientific Research were primarily directed toward the development of improved formal models for the methodologies for representation of vague and imprecise behavior known as possibilistic or "fuzzy" logic. We developed two classes of models, based on the notions of similarity and utility, respectively, and we studied the relations between each of those models and possibilistic logic. We also studied the relations between the two types of models.

Our study of similarity-based models sought a better understanding of the role of similarity-based structures in intelligent reasoning and of its applicability to analogical reasoning processes such as those employed, for example, in case-based reasoning.

Our investigations on utility-based models, on the other hand, sought a better understanding of the relation between possibilistic constructs and those of utility theory. A major goal of our research in utilitarian concepts was the development of a sound methodology to represent multiple objectives, their relations, and the explicit contextual circumstances that determine goal importance (e.g., "near an obstacle decrease importance of minimum-time goal"). In particular, we sought the development of "explainable" reasoning methods capable of tracing the inference path leading to a particular decision recommendation (e.g., why importance was given to some subgoals, how that assignment influenced choice).

Our research on approximate models led to the development of

- Enhanced models of similarity and utilitarian concepts, including improved understanding of notions such as joint and marginal similarity.
- Formal relationships between similarity-based notions and knowledge structures that permit the derivation of similarity measures—important in analogical reasoning methodologies such as case-based reasoning—from conditional rules.
- Enhanced fuzzy inference methods, based on these formal models.
- Methods for the control of systems under conditions of imprecision and uncertainty that blend *purposive* behaviors—oriented toward attainment of user-specified goals— and *reactive* behaviors, intended to cope with unexpected circumstances. These methods are capable of performing even when information about the operational environment is imprecise and ambiguous (*approximate maps*).

  Our methods were successfully applied to the control of autonomous mobile robots in experiments conducted at the AIC and at the First Robotics Competition of the American Association for Artificial Intelligence.
- Techniques to combine these control techniques based on numerical inference schemes with classical artificial intelligence (AI) planning procedures, which are primarily symbolic in nature.

- New, axiom-based techniques, for the dynamic resolution of conflicts between competing goals.
- Generalizations of dynamic programming techniques, based on concepts and structures from multivalued logic, that provide the bases for an evolving methodology for the automated generation of control and decision rules.

This document is composed of two major parts. The first part, introduced in this section, is a summary of the major research papers produced under the sponsored research. These papers constitute the second part.

Applicability of Methodology to DoD and the Commercial World

The concepts and techniques sought by our research program are basic in nature and are applicable to a wide variety of decision, control, and information-processing problems.

At the theoretical level, our research is directed toward the clarification of major issues in the nature of inconvenient, but nonetheless important, features of knowledge about complex systems: ambiguity, imprecision, uncertainty, and vagueness. In particular, we seek to clarify differences between aspects of information that result in imperfect knowledge about the state of the world (e.g., ignorance or uncertainty) and issues of applicability of knowledge items to specific situations. Problems caused by uncertain information are usually better dealt with by probabilistic reasoning methods. Examination of issues of knowledge suitability aspects requires, as shown in our earlier investigations, studies of notions of preference and similarity.

Utilitarian notions, such as cost or preference, are ubiquitous in the design and assessment of any practical system. Our approach, based on a sound, logic-based, treatment of preference and utility, has clarified numerous issues related to the specification of performance goals, notably the representation of relations and tradeoffs between multiple, possibly conflicting, objectives.

Our treatment of related similarity issues has led also to insights on analogical reasoning processes such as they are used, for example, in case-based reasoning.

At a more applied level, we have placed emphasis on evaluating the applicability of our theoretical results to important problems of DoD interest. In particular, we have sought to develop concepts that provide theoretical support for sound methods to treat complex decision and control problems.

The computational methods developed in our research have broad applicability to the regulation of a wide variety of control systems, to the planning of complex task activities, and to the support of decision makers dealing with large real-world systems. Of particular interest to DoD is the ability of our methodology to develop robust plans to guide manufacturing processes, strategic and logistic missions, and the actions of various autonomous vehicles.

We are particularly interested in the application of these techniques to the design and modification of complex systems (e.g., software systems, airplanes) where excessively detailed initial specifications lead to inflexible design processes and to large reengineering costs. This methodology may also be applied to control the acquisition of knowledge and the learning activities of intelligent systems.

2 Accomplishments

We briefly review here the major results of our research, which encompasses advances in the understanding of basic concepts and methods of possibilistic reasoning, developments in the planning and reactive control of systems operating under conditions of imprecision and uncertainty, studies of relationships between similarity measures and knowledge-based structures, and development of techniques for automated generation of control and decision rules.

Basic Research on Approximate Reasoning Methods

Approximate reasoning is a branch of AI concerned with the modeling and analysis of systems that are known under conditions of imprecision and uncertainty.

The ubiquitous presence of information that is partial, imprecise, and uncertain is a common characteristic of many important technological problems, including data sensing and fusion, situation assessment, decision support, device design, path planning, robotics, information retrieval, logistic planning, and fault diagnosis. In the military context, the importance of approximate reasoning is heightened by the need to process information having diverse levels of credibility, which is often distorted through the efforts of enemy agents.

Recent results by the author have established a unified framework for the description and comparison of diverse approximate reasoning methodologies.

This framework, which is based on the notion of *possible world*, has shown that basic conceptual differences exist between probabilistic methods—based on the concept of likelihood—and possibilistic methods—based on the related concepts of similarity and preference.

Approximate Reasoning and Possible Worlds

Possible worlds are the possible scenarios or situations describing the state of the real system or, in short, the possible, or logically conceivable, answers to a system analysis question.

In a classical reasoning problem, characterized by precise information, it is possible to provide answers that are not ambiguous,—i.e., that describe a system to the desired level of detail. In an approximate-reasoning problem, on the other hand, a question that has

either a "yes" or a "no" answer (e.g., whether a statement or proposition is true) may not be answered in unambiguous fashion, since the available information (or *evidence*) may be insufficient to infer the correct answer. In other words, there are scenarios where the answer may be yes, and scenarios where the answer may be no. The set of such scenarios, situations, or possible worlds is called the *evidential set*.

Existing approximate-reasoning methods may be said to be based on procedures that describe certain properties of such evidential sets. Two major classes of methods have been proposed, each leading to a different type of description. These approaches are called *probabilistic* and *possibilistic* reasoning, respectively.

Probabilistic Reasoning Methods

Probabilistic reasoning methods are concerned with the description of the likelihood that the answer to certain questions about the state of the system may be either "yes" or "no."

For example, if we are interested in the weather at a certain location and we ask the question "will it rain in three hours from now?," a probabilistic system may respond with an answer of the type "the probability that it will rain in three hours is 60%." Such a probability estimate may be the result of the examination of historical records of the weather under *similar* circumstances (*objective* probabilities), or the result of assessing the willingness of a weather expert to bet a certain amount of money in a gamble that pays off if, after three hours, it actually rains (*subjective* probabilities).

In some cases, the necessary records or elements of judgment required to derive such probability values may themselves be unavailable. Often, under such conditions we may say only that the probability of rain is between 60% and 80%. These *interval estimates* are the bases of so-called interval, or generalized, probability methods.

A major example of this type of method is the Dempster-Shafer calculus of evidence. This method may also be explained, using our unified framework, as the result of examining how often, in past similar experiences, it has been possible to answer questions of interest with a definite "yes." Since, in general, lack of information may sometimes prevent such a question from being answered with a firm "yes" or a firm "no," the sum of frequencies for such answers may be less than 1.

This simple idea—arising naturally from information that is both *uncertain* and *imprecise*—was formalized in a possible-worlds model based on the use of so-called *epistemic* logics.

Epistemic logics are formalisms that distinguish between the truth of a statement and the fact that such statement is *known* to be true to some reasoning agent. In other words, the functions of the calculus of evidence (called *mass distributions* and *support functions* by Shafer) measure the likelihood that a proposition may be known (or deduced) to be true if additional evidence were to be added to our imperfect information.

Possibilistic Reasoning Methods

In contrast with probabilistic reasoning, possibilistic reasoning describes the evidential set in terms of the resemblance of its "scenarios" to some reference situations. While probabilistic reasoning, concerned with proportions of truth or falsehood of propositions, finds a natural formalization on the notion of *set measure* (i.e., the relative proportion of "yes" versus "no" scenarios in the evidential set), possibilistic approaches find a ready expression using measures of similarity or resemblance between pairs of possible worlds.

A typical possibilistic scheme requires a rich variety of possible alternatives to a given question, so that questions such as "will it rain at least 1 inch in the next 24 hours?" may be answered by statements such as "I do not know, but I am sure that it will rain at least 3/4 of an inch" (where a probabilistic answer might have been "I do not know, but the chances of such precipitation are 60%").

Theoretical Research

During the reported research we continued to investigate the formal bases of possibilistic logic placing emphasis on interpretations of possibilistic constructs in terms of utility-oriented notions.

Our interest in concepts concerned with the utility of certain outcomes or the related notion of preference between alternatives stems from two major considerations.

At an informal level, it seems clear that, in problem-solving contexts, situations (i.e., possible worlds) are similar if there do not exist relevant preference considerations that render one of them considerably more desirable than the other. In other words, two situations resemble each other if they are nearly equally preferable from every important problem-relevant viewpoint.

From a purely formal—although hardly unrelated—perspective, preference functions are readily modeled by mathematical structures called *fuzzy preorders* (i.e., generalization of the mathematical notion of order relation), which have been shown to be closely related to the notion of *similarity* (which is, itself, a generalization of the classical notion of equivalence relation).

Exploiting the relations between metric (i.e., similarity, resemblance, distance) and utilitarian concepts, we developed utility-based interpretations that generalize the well-known interpretation of Bellman and Zadeh [1] that advanced the notion of fuzzy sets as "elastic constraints" on problem solutions.

Exploring comparisons between metric and utilitarian structures, we further developed the concepts of relative, marginal, conditional, and joint utility on the basis of their utilitarian counterparts. Some of these notions were mentioned in a recent technical paper [5], which is included in Part II of this report. A full treatment of a formal theory of similarity, and of its relations with analogical knowledge structures, is currently being prepared.

In our recent paper [5] we also improve the original interpretation of the basic inferential operation of fuzzy logic—the *generalized modus ponens*—by allowing unconditioned and conditional knowledge to be generated by different bodies of evidence. On the basis of these results and independent investigations, researchers at the Center of Advanced Studies in Blanes, with whom the principal investigator has collaborated and visited since the mid 1980s, were able to prove [4] that other important interpretations of fuzzy logic, such as the *possibilistic logic* of Dubois and Prade and *fuzzy-truth valued logic*, are formally subsumed by the similarity logic originally proposed by the principal investigator in 1991 [7].

Our semantic models have also led to new insights on the meaning of possibility distributions used to represent vague knowledge about real-world systems.

Although the meaning of possibility distributions as measures of the desirability of a particular state of affairs (e.g., the adequacy of a control policy or the overall utility of some outcome) has been understood for some time, the interpretations of vague statements about the world such as "The distance between obstacles is approximately 20 meters," and that of vague rules such as "If the temperature T is high, then the volume V is small," have remained largely unexplored.

Our studies at the conceptual level have now provided, however, a basic explanation of these declarative statements based on a utility-oriented interpretation of modeling. Informally, this interpretation considers the modeler, being interested in gaining knowledge about the state of the world, as an agent that must make decisions about the extent by which certain statements may be construed as correct descriptors of the state of a system or of its behavior. The errors made by assuming certain statements as facts or as applicable rules are quantified by possibility distributions. Should the modeler decide to employ a particular distribution, representing vague knowledge in his deliberations, he risks making erroneous assessments—the magnitude of which is numerically bound by the values of the distribution.

For example, should he decide to accept the validity of the vague proposition "John is tall," he is indicating his willingness to derive results that will be erroneous to the extent that the actual state of affairs matches the vague proposition. If John turns out to be 7 feet tall (a value assumed to have a possibility value of 1), then use of the statement "John is tall" will not result in substantial errors when reality is assessed. On the other hand, if John is only 5 feet tall (a value assumed to have a possibility of 0), then it is *possible* that the conclusions inferred from the model may be completely erroneous (note the emphasis on "may" as it is impossible to assert, in general, that a modeling error will *necessarily* lead to significant mistakes in hypothesis assessment).

From this perspective, vague models may be seen as statements of willingness, on the part of the modeler, to accept consequences, quantified by certain measures of potential loss, as the result of using inferences based on incorrect models of reality. In conventional cases, where all constraints are "crisp," such a disposition has the usual consequences—i.e., results can only be trusted if the assumptions agree with reality.

This interpretation of declarative statements in terms of potential modeling payoffs also

permits a clearer delineation of the role played by probabilistic constructs—measuring the tendency or propensity of nature to act in certain ways—as distinct entities from possibilistic structures—that assess the consequences of modeling errors. In our view, the two basic components of decision-making, namely utility and likelihood, are represented, in our semantic models, by formalisms intended to capture the properties of probabilistic concepts such as frequency or by theories dealing with the rational bases of preference (i.e., why one situation is deemed better than another).

It must be remarked, however, that although in most situations the interaction between probability and possibility will take the form of a probability distribution over utility values, in some cases that interaction has a more complex nature as when, for example, the vague fact "The obstacle is approximately 20 meters from point P" must be given an interpretation. This statement may be thought of as an assessment of the utility $\mathbf{Poss}(x)$ of assuming that the obstacle is at a distance of 20 meters from P when its actual distance is x, or, alternatively, as a declaration that the probability of finding the object at a distance x is given by $\mathbf{Poss}(x)$ (i.e., the utility value is given directly by the probability of risk).

Our investigations also resulted in new methods for the combination of utility functions. Among these are techniques that permit the construction of complex utility functions using (generalized) logical operators. These methods permit the description of the relative importance of goals, relations between operational contexts and objectives, and criteria to be used to resolve dynamic conflicts between competing objectives.

Unlike other approaches, which mainly stress weighted combinations of numerical indexes of performance, our logic-based techniques facilitate the explanation of the rationale leading to control and decision recommendations.

The ability to explicitly limit certain actions (e.g., forcing compliant motion in a manipulator) combined with the gradual nature of the underlying logics (based on a continuous truth scale) permits the easy development and refinement of control and decision systems that smoothly combine actions seeking the attainment of multiple, possibly conflicting, goals, while responding to unforeseen environmental circumstances.

We are currently continuing to expand and improve these methods, focusing particularly on relationships with multiattribute utility theory [6].

Closely related to these results is our derivation of a new technique for resolution of conflicts between competing goals. This method, based on an axiomatic approach, defines acceptable outcomes of any conflict-resolution algorithm by means of a system of axioms that captures the notion of constraint relaxation. On the basis of the analysis of these axiomatic constraints, it is possible to derive an expression that permits the calculation of a (fuzzy) set of "acceptable" choices. Unlike other approaches, this scheme is based on a partial order of the problem goals and does not require specification of values representing relative importance of the problem objectives.

We have also initiated investigations leading to the development of efficient logic programming methods for the specification of approximate models. Our first step is the formal

specification of a logic incorporating possibilistic and probabilistic notions. We are currently engaged, with Dr. Alessandro Saffiotti of the Free University of Brussels (who visited SRI as an International Fellow from 1992 to 1993), in such a task, which includes specification of syntax, semantics, and, most important, proof procedures for such a complex multivalued logic. Our point of departure is, once again, our formal semantic models. Our studies have been helped considerably by the previously mentioned results of Esteva , García and Godó [4], which have further strengthened knowledge about the meaning of possibilistic structures.

Control in Uncertain Environments

The interaction between humans and their environment requires the development of effective techniques to control the behavior of a wide variety of real-world systems. The continued growth of control theory and its application to a variety of problems, from process regulation and production of logistic and strategic plans to the design and modification of complex assemblies, simply reflects the importance attached to methods that direct the evolution of complex processes and systems.

In many planning and control applications there exist requirements that go beyond the mere production of a plan under certain assumptions about the nature of a system. In these problems, the underlying systems operate under conditions of uncertainty that make impossible the prediction of changes to the environment where the system operates. Under these conditions, any policy without sufficient flexibility is bound to be of little use as, typically, dynamic environmental changes prevent further utilization of the policies recommended by the plan.

Our principal research objective in the area of control and decision-support systems was the application of our theoretical results to the development of robust controllers. We sought methods to build controllers capable of reacting in a flexible, adaptive fashion to changes in the environment while still continuing, to the best possible extent, to try to achieve explicit control goals. We also investigated the generation of control and decision strategies that respond in a gradual, smooth, fashion to perceived modifications in environmental conditions.

If a dynamic control problem is thought of as the regulation of the behavior of a system, then we may classify such behavior as being essentially *purposive* (e.g., transport troops and equipment to some place in an efficient fashion) or *reactive* (e.g., overcome a temporary difficulty such as troublesome weather). In this view, the development of a control policy is considered to be equivalent to the identification of purposive and reactive behaviors, and of procedures to activate, deactivate, and integrate (or "blend") them.

Each of the just-mentioned components of a control policy is required to assure successful performance of a robust controller. Purposive behaviors are obviously required to assure that the system attains its goals. Reactive behaviors, on the other hand, are intended to provide responsiveness to a myriad of possible but difficult-to-predict environmental changes. Finally, adequate activation and blending techniques are needed to assure smooth transition

between behaviors and to permit partial attainment of goals while responding to evolving circumstances.

The requirement to trade off a degree of goal attainment with responsiveness is essential to assure that incompatible goals, sought by various possible modes of operational behavior, may be successfully integrated. In a typical logistic planning problem, for example, demands to deliver equipment at a certain time may be traded off with requirements to attain a minimum level of security (i.e., within bounds, delays may be permitted if they decrease mission risk).

Control/Decision Systems

Our investigations resulted in the development of a general system control methodology for the dynamic management of multiple, possibly conflicting, objectives under conditions of imprecision and uncertainty. This technology is based on methods that compute numerical *desirability measures* for control actions as assessed from the viewpoint of specific goals or objectives. These individual measures may then be combined, using explicit descriptions of rule applicability, goal importance, and allowable tradeoffs, into combined measures that gauge the overall suitability of each possible control action.

Our approach relies on dynamic management of a set of control rules on the basis of changes in the operational context of the plant. This approach is based on our theoretical results, which interpret fuzzy logic as a generalized inferential methodology placing emphasis on the measurement of the utility of various states of affairs (e.g., the desirability of utilizing some amount of a resource, given the current circumstances). We also have successfully combined numerical possibilistic techniques with conventional AI planning techniques into an integrated methodology for the development of flexible plans to attain multiple system goals.

These developments have extended the applicability both of conventional planning techniques and of approaches to represent and manipulate "elastic" constraints—that is, constraints that may be satisfied to various degrees. Related methods also allow the explicit specification and manipulation of multiple, possibly conflicting, goals and the representation of context-dependent tradeoff policies.

The desirability measures used to determine control and decision choices are computed by application of control rules associated with a particular goal or objective. This knowledge base contains either explicit rules of the form

"If the state is in set S, then the control must be in set C,"

or, in more complex cases, rules that are capable, likewise, of inferring (through a sequence of steps) the suitability of each potential control action. For example, rules to drive a car to some intersection measure the desirability of each steering and accelerating action as a function of the perceived position of the car.

A well-known generalization of the deductive rule of *modus ponens* is employed to perform the inferential steps. In the generalized modus ponens, each rule fires to a different degree, depending on the degree of matching of the rule antecedent and the present situation. If the match is exact (i.e., the current state implies the truth of the antecedent), then the consequent is asserted. If the match is not exact, then a more general conclusion is derived. When the match is very poor, the outcome is noninformative (i.e., for all we know, any control might work).

Desirability functions are associated both with purposive goals and with reactive behaviors. In an autonomous vehicle application, for example, a purposive goal may be to reach a location or to observe some landmark, while a reactive behavior may be the result of following procedures for obstacle avoidance.

The computational mechanism utilized to determine the importance of activation or deactivation of behaviors and to specify related desirability functions is called a *control structure*. A control structure specifies the contextual conditions (in terms of requirements for certain events or structures to be perceived or observed by the system) that must be met for an associated behavior to be activated. The rules associated with such a behavior are then used to determine the suitability of individual actions.

Since contexts are typically activated in a gradual fashion (e.g., rules to turn a vehicle are not turned instantly on or off, but they are gradually activated as a function of the distance between the current position of a vehicle and an ideal turning point), several behaviors may be present at any time, each activated to a different degree. The result of this aggregation of behaviors has been shown experimentally to lead to smooth transition between behaviors and to high responsiveness to unforeseen events.

Our methodology does not require precise, certain knowledge of the characteristics of the system and of its environment. In vehicle control applications, for example, we have relied on approximate maps of the workspace. These maps consist primarily of connectivity information, describing the general layout of the workspace. This information is complemented by approximate metric information characterizing important features of the environment (e.g., approximate distances between objects, approximate angles of intersection between hallways).

In the course of our research we have applied these techniques to the control of the local motion of an autonomous vehicle. We have simulated and tested their approach in SRI's autonomous vehicle in experiments conducted in our AIC and in the context of the First Robotics Competition of the American Association for Artificial Intelligence. SRI's autonomous vehicle, which was one of only two that did not rely on instrumentation (i.e., bar code markers, emitters) of the competition area, was commended by the judges for its ability to quickly and smoothly react to unforeseen circumstances.

The experience gained in our experiments provided, in turn, additional bases for our theoretical research. On such bases we have developed a theory of reactive control [8], which

we have documented in a technical report and in a forthcoming paper to be published in *Artificial Intelligence.*

It is important to remark that although our techniques were tested in an autonomous mobile-device application, these methods have wide applicability to a number of decision and control problems.

Results of our research efforts are given in detail in the publications Part II of this report. A videotape showing the performance of SRI's autonomous vehicle under the guidance of a fuzzy controller based on our results is also available from the AIC.

In addition to our experiments with autonomous mobile platforms, we are currently engaged in the development of controllers for a number of systems of practical importance. Among these, two specific platforms deserve special mention.

The first of these systems is a prototype of a lightweight, flexible manipulator, which must be moved subject to constraints that limit both its motion and the extent of motion-induced vibrations on the manipulator's arm. The second prototype is provided by a complex model of a hybrid gasoline-electric automobile power-generation plant that must be controlled so as to attain a number of objectives in response to unpredictable driver demands.

A noteworthy characteristic of our approach to the solution of these control problems is its reliance on the development of approximate, qualitative models, capable of being validated as true descriptions of the actual real-world systems that they model. Our techniques, based on the theoretical results discussed above, are based on sound procedures for the analysis of these models and for the deductive derivation of controllers. This approach is significantly different from most of the heuristic and intuitive techniques that have, so far, characterized the application of fuzzy-logic ideas.

Planning

The need for robust, flexible systems extends well beyond needs for the local control and regulation of devices. A wide variety of important applications ranging from logistics and flexible manufacturing to complex system design and modification (e.g., software systems) would greatly benefit from methods to determine decision policies that may be readily modified because of unforeseen circumstances.

Conventional AI planners do not have, at this time, capabilities for the representation of elastic constraints, for their manipulation, or for quick modification of plans in response to variations in initial assumptions. As is the case for control problems, the difficulties center on the inability to efficiently determine the performance effects of variations in the control and decision parameters.

On the other hand, conventional AI planners, relying primarily on symbolic techniques, are well suited to identifying plans to attain multiple objectives on the basis of effective searches of the planning space and careful determination of the relations between goals.

In particular, conventional AI planners are capable of identifying major feasible strategies, specific subgoals to be achieved, and temporal sequencing conditions (e.g., do not initiate Task B until Task A is completed).

Consideration of the complementary capabilities of symbolic planning techniques with numerical inference methods suggests that these technologies might be effectively combined into an integrated approach that incorporates the strong points of each methodology. In particular, it will be desirable to take advantage of the ability of symbolic techniques to break down complex planning problems into separate subproblems and that of numeric approaches to specify desirability of a number of alternative control actions.

We investigated the combination of techniques based on the notions of control structure and desirability measure with classical AI planning approaches, such as those that are incorporated in SRI's planning systems SIPE and PRS, to produce a general methodology to the production of flexible planning and control policies.

The overall approach employed by this methodology relies on the determination, at the high level, of subgoals to be attained by the control policy. These subgoals, however, are elastic objectives in the sense that they may be attained to different degrees. While, ideally, the controller would like to attain each subgoal to the maximum possible extent, inherent conflicts between goals combined with requirements for reactivity to unforeseen circumstances may result in some of them being achieved at lesser degrees. For example, a manufacturing plan may require certain use of resources for optimal productivity. This usage may be dynamically modified, however, in response to circumstances (e.g., slowdowns in a station or buffer overflow), and such changes may result in lower, yet acceptable, yields. In this way, the plan may be effectively modified without undergoing expensive and time-consuming replanning operations.

This methodological approach seeks to emulate the human capability to formulate plans as general, high-level, descriptions of the overall strategy to be followed to attain some goals, which are specified in terms that are flexible enough to permit real-time adjustments during their execution. For example, a flight plan may be specified in terms of various points to be reached and areas to be avoided. However, these milestones and constraints are, with few exceptions, to be interpreted as ideal, rather than actual desiderata. Similarly, instructions to park a car are usually meant to describe prototypical maneuvers that may be followed approximately and changed, during execution, according to the circumstances.

The high-level planner may also be thought of as a mechanism that identifies the rules to be followed by the controller during execution of the plan—that is, the numerical standards that will be used to judge controller performance. The desirability functions that embody those measures are usually not explicitly given but, rather, they are computed by application of the inferential mechanisms of fuzzy logic to a knowledge base of rules that promote performance of certain behaviors.

Experiments on the integration of these control techniques with conventional planning approaches using both a goal-regression planner and a procedural reasoning system were

completed in early 1993. These experiments, presented in detail in the publications in Part II, have established the feasibility of this two-level approach to planning and control where a high-level process specifies partial, flexible, plans, and a low-level process carries those specifications and reacts to environmental changes.

Similarity-based Structures and Analogical Reasoning

We have further studied the relation between possibilistic methods and similarity-based techniques seeking to further understand the relations that exist between logic concepts (predicates, variables) and metric structures (joint, marginal, conditional similarities) with a view to the development of analogical reasoning methods (e.g., case-based reasoning). The results of these examinations have provided a logical basis for the constructive development of similarity measures on the basis of domain knowledge expressed by vague facts and conditional knowledge rules. We have also benefitted from insights on the relations between metric and utilitarian interpretations that have permitted incorporation of technical knowledge on the structure of utility functions into the study of analogical-reasoning issues.

We have obtained new results linking joint and conditional similarities that have provided the bases for current studies on the identification of conditioning formulas: the backbone of any evidence-combination approach. Results of this research have been published in a recent paper, included in Part II, which is coauthored by colleagues of the principal investigator at the Center of Advanced Studies, Blanes, and at the University of the Balearic Islands, Spain. We expect to continue these joint studies, focusing on alternative interpretations of the notion of conditional knowledge and its determination.

We have also developed mechanisms to translate statements describing (typically in a multivalued logic) facts about the world into actual measures of similarity between different situations (i.e., possible worlds). This research is based on our previous models relating the notions of possibility and similarity.

Control/Decision Rule Generation

The success of any multistage sequential decision scheme for the planning and control of complex real-world systems hinges on the ability to successfully relate overall goals (e.g., minimal time to reach some state) with specific instantaneous actions that promote those goals.

In our approach to control, instantaneous actions are controlled by desirability functions that gauge the relative quality of potential decisions as promoters of some specific goal or objective. Overall goals or objectives are usually expressed as elastic constraints on system behavior.

In our methodology, desirability functions are the output of inferential processes operating on a dynamic knowledge base of control rules. In our experiments so far, the actual rules

have been derived either as the result of immediate analysis of the particular system at hand or as the result of trial-and-error refinement of heuristic rule sets.

One of the most important objectives of our continuing research is, however, the systematization and normalization of the control-synthesis process by

1. Representation of the underlying system by a qualitative model based upon sound logical constructs
2. Description of system goals and constraints by means of the same constructs
3. Joint analysis of system characteristics and problem constraints by sound inferential methods that derive the desired control and decision strategies
4. Determination, again by application of sound deductive procedures, of properties of the system being regulated

The research reported in this document has attained these objectives while providing strong semantic bases (i.e., similarity and utility interpretations) for the development of further representation and analysis methods.

Seeking to develop methods for the attainment of our remaining goals, we turned our attention toward methodologies that promote derivation of control strategies by deductive analysis of system models, constraints, and objectives. Because of its generality, simplicity, and adaptability to multivalued-logic reformulation we concentrated on dynamic programming.

Dynamic programming is the most powerful method for the synthesis of state-based decision strategies and feedback controllers, being, more generally, an extremely powerful procedure for treating a variety of optimization problems.

Unfortunately, the usefulness of this methodology is limited by the need to store solutions as large tables, even for problems of relatively low dimensionality. A number of recent theoretical (approximation theorems) and practical (fuzzy-controller design) results [9] have shown, however, that solutions of dynamic programming problems may be efficiently approximated by compact sets of fuzzy rules. Furthermore, a number of efficient techniques, based primarily on clustering and neural-network methods, have been developed and successfully applied [2, 3] to derive such approximations.

Our research on multistage decision processes has resulted in a logic-based reformulation of the Hamilton-Jacobi-Bellman (HJB) equation as a fixed-point logical expression. The essential observation leading to this reformulation is that, when dynamic behavior and performance objectives are modeled by possibility distributions according to the ideas first advanced by Bellman and Zadeh [1], the principle of optimality of Bellman may be rewritten in logical form as the fixed-point logical equation

$$\texttt{Good}(x) \Leftrightarrow \exists y, u\ \texttt{Feasible(x,y,u)} \wedge \texttt{Efficient(x,y,u)} \wedge \texttt{Good}(y)\,,$$

where `Good` is a numerical measure of the quality of the *optimal* set of decisions guiding a system toward achievement of dynamic goals, `Feasible` is a multivalued-logic predicate measuring the physical feasibility of moving the state from x to y using the control u, and `Efficient` is another multivalued-logic predicate measuring the efficiency of such a movement. In simpler worlds, this fixed-point expression simply states that there exists an optimal strategy leading the system from x toward its goals if there exists an optimal subtrajectory leading x to some y and if there exists an optimal strategy from y toward those objectives.

This reformulation of the HJB equation has led, in turn, to a generalization that does not define quality of a control strategy in terms of its optimality, measuring, instead, the degree of compliance of any strategy (i.e., a feedback function) with prototypical strategies known to produce adequate behavior. The essential novel element of that formulation is the explicit introduction of expressions that measure compliance with specific strategies, thus relaxing the notion of adequacy as optimality that is built into classical dynamic programming formulations.

This generalization of the HJB equation is helpful as a source of insight into the solution of a wide variety of operational research problems and their generalizations. Current research based on these ideas is aimed at three major objectives:

1. The iterative derivation of solutions for the generalized equation by means of algorithms that improve approximations of the feedback function by fuzzy rules
2. Identification of relations between solutions of the HJB equation and their generalizations that provide insight leading to compact, effective approximations
3. Development of techniques that promote attainment of multiple, possibly conflicting, objectives, by dynamic relaxation and blending of feedback strategies

Following the successful representation of dynamic programming approaches for terminal-set control, using a utility-based, multivalued-logic representation, we have started to develop a series of recursive control approaches based on direct modeling of the adequacy of control schemes, using approaches that directly generalize the Hamilton-Jacobi-Bellman equation. Informally, the idea is that a "good" control steering a discrete-time system from a point P to a terminald -set is one that

1. Is economical in its first step
2. Results in a state, after the first step, that may be guided in a satisfactory fashion to the terminal set

When such statements are formalized using multivalued logic tools, the resulting fixed-point equation may be used to determine the degree of adequacy (as measured from a variety of viewpoints) of various control policies.

We are also investigating supervised learning methods to generate rules for the representation of performance values (as a function of state) and for the generation of the desirability measures.

3 Summary of Accomplishments of Principal Investigator

Research Publications

E.H. Ruspini, On the Semantics of Fuzzy Logic, *Int.J. Approximate Reasoning*, 5: pp. 45–88, 1991.

E.H. Ruspini, Approximate Reasoning: Past, Present, Future, *Information Sciences*, 57-58: 297–317, 1991.

E.H. Ruspini, J. Lowrance, and T. Strat, Understanding Evidential Reasoning, *Int. J. Approximate Reasoning*, 6: 401-425, 1992.

E.H. Ruspini, On Truth, Utility and Similarity. In *Proceedings of the 1991 International Fuzzy Engineering Symposium*, Yokohama, Japan, 1991.

E.H. Ruspini, Control of Autonomous Vehicles using Fuzzy Logic. In L. Valverde, editor, *Proceedings of the 1992 International Conference on the Management and Processing of Uncertainty by Expert Systems*, Palma, Spain, 1992.

E.H. Ruspini, Autonomous Vehicles, Approximate Maps, and Fuzzy Logic. In *Proceedings of the 1992 NASA Workshop on Fuzzy Logic and Neural Networks*, Houston, Texas, 1992.

E.H. Ruspini and D.C. Ruspini Autonomous Vehicle Motion Control using Fuzzy Logic. In *Proceedings IEEE Round Table on Fuzzy Logic, Neural Networks, and Intelligent Vehicles,* Tokyo, Japan, 1991.

E. H. Ruspini, Progress in Research on Autonomous Vehicle Motion Planning, In J. Yen and R. Langari (editors), *Industrial Applications of Fuzzy Logic*, forthcoming, 1994.

A. Saffiotti, E.H. Ruspini, and K. Konolige. Blending Reactivity and Goal-directedness in a Fuzzy Controller. *Proceedings of the Second IEEE International Conference on Fuzzy Systems,* San Francisco, California, 134–139, 1993.

C. Congdon, M. Huber, D. Kortenkamp, K. Konolige, K. Myers, E.H. Ruspini, and A. Saffiotti. CARMEL vs. Flakey: A comparison of two winners. *AI Magazine,* 14(1):49–57, Spring 1993.

C. Congdon, M. Huber, D. Kortenkamp, C. Bidlack, C. Cohen, S. Huffman, F. Koss, U. Raschke, T. Weymuth, K. Konolige, K. Myers, A. Saffiotti, E.H. Ruspini, and D. Musto. CARMEL *vs. Flakey: A comparison of two winners.* AAAI Technical Report RC-92-01, Menlo Park, California, 1994 Extended version of the paper published in *AI Magazine.*

A. Saffiotti, E.H. Ruspini, and K. Konolige. Robust Execution of Robot Plans using Fuzzy Logic. *Proceedings IJCAI-93 Workshop on Fuzzy Logic in Artificial Intelligence,* Chambery, France, August 1993. Also in A.L. Ralescu, editor, *Fuzzy Logic in Artificial Intelligence,* Lecture Notes in Artificial Intelligence No. 847, 24–37, Springer-Verlag, Berlin, 1994.

A. Saffiotti, E.H. Ruspini, and K. Konolige. A Fuzzy Controller for Flakey, an Autonomous Mobile Robot. *Proceedings of the Fuzzy Logic '93 Conference,* Burlingame, California, July 1993. To be published also in a forthcoming volume by Springer Verlag. (Extended version published as Technical Note No. 529, Artificial Intelligence Center, SRI International, Menlo Park, California, August 1993.)

A. Saffiotti, E.H. Ruspini, and K. Konolige. Using Fuzzy Logic for Autonomous Vehicle Planning. *Proceedings of the 1993 NAFIPS Conference,* Allentown, Pennsylvania, August 1993. To be published also in a forthcoming volume of selected papers.

A. Saffiotti, E.H. Ruspini, and K. Konolige. Robust Control of a Mobile Robot using Fuzzy Logic. *Proceedings of First European Congress on Fuzzy and Intelligent Technologies: EU-FIT '93* Aachen, Germany, September 1993.

A. Saffiotti, K. Konolige, and E.H. Ruspini. *A Multivalued Logic Approach to Integrating Planning and Control.* Technical Note No. 533, Artificial Intelligence Center, SRI International, Menlo Park, California, June 1993.

F. Esteva, P. García, L. Godó, E.H. Ruspini, and L. Valverde On Similarity Logic and the Generalized Modus Ponens. *Proceedings of the 1994 IEEE International Conference on Fuzzy Systems,* 1423–1427, Orlando, Florida, June 1994.

A. Saffiotti, K. Konolige, and E. Ruspini. A Multivalued Logic Approach to Integrating Planning and Control. *Artificial Intelligence,* forthcoming, 1995.

A. Saffiotti, N. Helft, K. Konolige, J. Lowrance, K. Myers, D. Musto, E.H. Ruspini, and L. Wesley A Fuzzy Controller for Flakey, the Robot (Video and Video Abstract), *Proceedings of the Eleventh National Conference on Artificial Intelligence,* AAAI and MIT Press, Washington, DC 1993.

E. H. Ruspini. Fuzzy Logic-Based Planning and Reactive Control of Autonomous Mobile Robots. In *Proceedings of the 1995 IEEE/IFES International Conference on Fuzzy Systems,* Yokohama, Japan, forthcoming, 1995 (Invited plenary lecture).

E.H. Ruspini. The Semantic Bases of Fuzzy Logic. In C. H. Chen, editor, *Fuzzy Logic and Neural Network Handbook*, McGraw-Hill, forthcoming, 1995.

E.H. Ruspini. Reactive Control using Fuzzy Logic. In C. H. Chen, editor, *Fuzzy Logic and Neural Network Handbook*, McGraw-Hill, forthcoming, 1995.

E.H. Ruspini. Fuzzy Sets, Relations, and Fuzzy-set Calculus. In J. David Irwin, editor, *The Industrial Electronics Handbook,* CRC Press/IEEE Press, forthcoming, 1995.

E.H. Ruspini. Reactive Fuzzy Controllers. In K. Hirota, editor, *Advances in Fuzzy Systems: Applications and Theory— Volume 2: Industrial Applications of Fuzzy Technology in the World*, World Scientific Publishing Co., forthcoming, 1995.

E.H. Ruspini, P.P. Bonissone, and L. Valverde. *Fuzzy Logic.* Textbook (in preparation, expected publication in Summer 1995).

E.H. Ruspini, P.P. Bonissone, and W. Pedrycz, editors. *Handbook of Fuzzy Computation.* Institute of Physics and Oxford University Press (in preparation, expected publication in May 1996).

Postdoctoral associates supported

Dr. Alessandro Saffiotti and Dr. Daniella Musto, supported by the Italian National Research Council, visited the Artificial Intelligence Center and engaged in research on Approximate Reasoning under the direction of Dr. Enrique H. Ruspini.

Dr. Nicolas Helft, under sponsorship from SRI, collaborated with Dr. Ruspini, during part of the reporting period, in investigations on the application of fuzzy-logic methods to planning and control.

Special Honors and Recognitions

SRI Institute Fellow, 1992.

Visiting Scholar, NASA Ames Research Center, July 1994-February 1995.

General Chairman, 1993 IEEE International Conference on Neural Networks (ICNN '93).

General Chairman, Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '93).

Tutorials Chairman, First IEEE International Conference on Fuzzy Systems (FUZZ- IEEE'92), San Diego, California, 1992.

Invited Speaker, 1992 International Conference on Processing and Management of Uncertainty, Palma, Spain, 1992.

Invited Panelist, 1992 IEEE International Conference on Intelligent Robots and Systems, Raleigh, North Carolina, 1992.

Invited Panelist, 1992 Conference on Uncertainty in Artificial Intelligence, Stanford, California, 1992.

Invited Tutorial Speaker, First IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'92), San Diego, California, 1992.

Invited Speaker, Boeing Workshop on Fuzzy Logic, Seattle, Washington, December 1991.

Invited Speaker, IEEE Phoenix FUZZ-tival, Phoenix, Arizona, 1992.

Technical Program Cochair, Third IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '94).

Program Chairman, 1994 Jornadas Argentinas de Informática e Investigación Operativa.

Other Achievements

Member of the Editorial Board of

- IEEE Transactions on Fuzzy Systems
  - (Associate Editor 1992-93, Member of the Advisory Board 1993-1994)
- International Journal of Approximate Reasoning (Associate Editor)
- IEEE Transactions on Fuzzy Systems (Associate Editor)
- Journal of Applied Nonstandard Logics
- International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems

Member of the Technical Group 2.6 *Database and Knowledge Representation*, IFIP (International Federation of Information Processing Societies).

Editor-in-Chief, *Handbook of Fuzzy Computation*, to be published by the Institute of Physics and Oxford University Press.

Member of the Program Committee of:

- 1994 International Conference on the Processing and Management of Uncertainty
- 1992 International Conference on the Processing and Management of Uncertainty
- 1992 IEEE International Conference on Fuzzy Systems
- 1993 IEEE International Conference on Fuzzy Systems
- 1994 IEEE International Conference on Fuzzy Systems
- 1995 IEEE/IFES International Conference on Fuzzy Systems

1994 AAAI National Conference on Artificial Intelligence
1992 Conference on Industrial Applications of Fuzzy Systems, College Station, Texas
1993 Conference on Industrial Applications of Fuzzy Systems, College Station, Texas
1994 International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan
1992 International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan
1993 European Congress on Fuzzy and Intelligent Technologies, Aachen, Germany
1994 European Congress on Fuzzy and Intelligent Technologies, Aachen, Germany
1992 Annual Conference on Uncertainty in Artificial Intelligence
1993 Annual Conference on Uncertainty in Artificial Intelligence

Member, Technical Group 2.6 (Databases), International Federation of Information Processing Societies (IFIP), 1982–1994.

During the period of this research, Dr. Enrique H. Ruspini also delivered invited lectures at Stanford University, Oxford University, University of California (Berkeley), University of Paris, University of Erlangen (Germany), National Technological University, San Jose State University, Third Dortmund Fuzzy Days (Dortmund, Germany), University of the Balearic Islands (Spain), University of Valladolid (Spain), University of Oviedo (Spain) University of Buenos Aires (Argentina), and University of La Plata (Argentina).

In addition, Dr. Ruspini recorded two video lectures under the auspices of the IEEE Educational Activities Board and produced, together with A. Bergman, another IEEE-sponsored video discussing the state of the art in fuzzy logic and neural networks.

References

[1] R.E. Bellman and L.A. Zadeh. Decision-making in a Fuzzy Environment. *Management Sciences,* 17: 141–164, 1966.

[2] H. Berenji, Y-Y. Chen, C-C. Lee, J-S. Jang, and S. Muruguesan. A Hierarchical Approach to Designing Approximate Reasoning-based Controllers for Dynamic Physical Systems. *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence,* Artificial Intelligence, 1990.

[3] H. Berenji. Fuzzy Logic Controllers. In R.R. Yager and L.A. Zadeh, editors, *An Introduction to Fuzzy Logic Applications in Intelligent Systems,* Kluwer Academic, Boston, 1991.

[4] F. Esteva, P. García-Calvés, and L. Godó. Relating and Extending Semantical Approaches to Possibilistic Reasoning. *International Journal of Approximate Reasoning,* 1994.

[5] F. Esteva, P. García-Calvés, L. Godó, E.H. Ruspini, and L. Valverde. On Similarity Logic and the Generalized Modus Ponens. In *Proceedings of the Third IEEE International Conference on Fuzzy Systems*, 1423–1427, Orlando, Florida, 1994.

[6] R. Keeney and H. Raiffa. Multiobjective Decision Making with Multiple Attributes. Wiley, New York, 19 .

[7] E.H. Ruspini. On the Semantics of Fuzzy Logic. *Int.J. Approximate Reasoning*, 5:45–88, 1991.

[8] A. Saffiotti, K. Konolige, and E. Ruspini. A Multivalued Logic Approach to Integrating Planning and Control. *Artificial Intelligence,* forthcoming, 1995.

[9] L.X. Wang. Fuzzy Systems are Universal Approximators. *Proceedings of the First IEEE International Conference on Fuzzy Systems*, 1163–1170, San Diego, California, 1992.

PART II

PUBLICATIONS

On the Semantics of Fuzzy Logic

On the Semantics of Fuzzy Logic*

Enrique H. Ruspini
SRI International, Menlo Park, California

ABSTRACT

This paper presents a formal characterization of the major concepts and constructs of fuzzy logic in terms of notions of distance, closeness, and similarity between pairs of possible worlds. The formalism is a direct extension (by recognition of multiple degrees of accessibility, conceivability, or reachability) of the major modal logic concepts of possible and necessary truth.

Given a function that maps pairs of possible worlds into a number between 0 and 1, generalizing the conventional concept of an equivalence relation, the major constructs of fuzzy logic (conditional and unconditioned possibility distributions) are defined in terms of this similarity relation using familiar concepts from the mathematical theory of metric spaces. This interpretation is different in nature and character from the typical, chance-oriented, meanings associated with probabilistic concepts, which are grounded on the mathematical notion of set measure. The similarity structure defines a topological notion of continuity in the space of possible worlds (and in that of its subsets, i.e., propositions) that allows a form of logical "extrapolation" between possible worlds.

This logical extrapolation operation corresponds to the major deductive rule of fuzzy logic – the compositional rule of inference or generalized modus ponens of Zadeh – an inferential operation that generalizes its classical counterpart by virtue of its ability to be utilized when propositions representing available evidence match only approximately the antecedents of conditional propositions. The relations between the similarity-based interpretation of the role of conditional possibility distributions and the approximate inferential procedures of Baldwin are also discussed.

A straightforward extension of the theory to the case where the similarity scale is symbolic rather than numeric is described. The problem of generating similarity functions from a given set of possibility distributions, with the latter interpreted as defining a number of (graded) discernibility relations and the former as the result

---

*To my friends Nadal Batle, Francesc Esteva, Ramón López de Mántaras, Enric Trillas, and Llorenç Valverde.

Address correspondence to Enrique H. Ruspini, AI Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025.

International Journal of Approximate Reasoning 1991; 5:45–88

655 Avenue of the Americas, New York, NY 10010 0888-613X/91/$3.50

of combining them into a joint measure of distinguishability between possible worlds, is briefly discussed.

KEYWORDS: ***fuzzy logic, semantics, modal logics, possible worlds, generalized modus ponens***

INTRODUCTION

This paper presents a semantic characterization of the major concepts and constructs of fuzzy logic in terms of notions of similarity, closeness, and proximity between possible states of a system that is being reasoned about. Informally, a "possible state" (to be formalized later using the notion of "possible world") is an assignment of a well-defined truth value (i.e., either *true* or *false*) to all relevant declarative knowledge statements about that system.

The primary goal that guided the research leading to the results presented in this work was one of conceptual clarification. A great deal of energy has been directed in the past few years to debating the methodological necessity and relative merits of various approximate reasoning methodologies. As a result of these exchanges, the need to consider certain nonclassical approaches has been questioned on a variety of bases.

Recognizing the need for the development of sound semantic formalisms that shed light on the nature of different approaches, I have pursued, in the past few years, a line of theoretical research seeking to describe various approximate reasoning methodologies using a common framework. These investigations have recently shown the close connection between the Dempster–Shafer [38] calculus of evidence [35] and epistemic logics. This relationship was elucidated by straightforward application of conventional probabilistic concepts to models of knowledge states that distinguish between the true of a proposition and knowledge (by rational agents) of that truth. Central to this development is the notion of "possible world" used by Carnap [6] to develop logical bases for probability theory.

The central notion of possible state of affairs is also the conceptual basis of the results presented in this paper, which is aimed at establishing the semantic bases of possibilistic logic with emphasis on the study of its possible relations and differences, if any, with probabilistic reasoning.

The results of this investigation clearly show that possibilistic logic can be interpreted in terms of nonprobabilistic concepts that are related to the notions of continuity and proximity. The major functional structures of fuzzy logic, possibility and necessity distributions,[1] may be defined in terms of the more primitive notion of similarity between possible states of a system using constructs that are the direct extension of well-known concepts in the theory of

[1] It is important to remark that the scope of this work is limited to the most fundamental concepts and constructs of fuzzy logic without examining related notions such as generalized quantifiers.

metric spaces. The topological metric structure that is so defined may be used to derive a sound inferential rule that is a form of logical "extrapolation." This rule is also shown to be the compositional rule of inference or generalized modus ponens proposed by Zadeh [53]. Conversely, possibility distributions—expressing resemblance in some specific regard—may be used to derive the actual similarity functions, discerning between possible worlds from the multiple points of view.

The constructs that are used to derive the interpretation presented in this paper are formally, structurally, and conceptually different from those that explain probabilistic reasoning, in either its objective or subjective interpretations, irrespective of methodological reliance on interval-based approaches to represent ignorance. The latter class of methods—measuring the relative proportion of the (either observed or believed) occurrence of some event—are based on the mathematical notion of set measure, while the former—seeking to establish similarities between situations that may be used for analogical reasoning—are related to the theory of distances and metric spaces.

This presentation of the relationships between similarity-based concepts and possibilistic notions, while grounded on a formal treatment that is based on rigorous logical and mathematical formalisms, will be kept at a level that is as informal as possible. The purpose of this presentation style is to facilitate comprehension of major ideas without the clutter that would otherwise need to be introduced keep matters strictly precise. For this reason, I will refrain from formal introduction of structures and axiom schemata, that, although correct and proper, may encumber understanding of the basic concepts.

Before we proceed to the detailed consideration of semantic models, I must briefly remark on the epistemological implication of these developments. The present interpretation is not the only that may be advanced to define the notion of possibility in terms of simpler concepts, nor do I claim that it may not be sometimes possible, even desirable, to model possibilistic structures from other bases. My intent is not to prove the conceptual superiority of one approach over another or to argue about the relative utility of different technologies. Rather, I hope that these results have contributed to establish the basic conceptual differences in the treatment of imprecise and uncertain information that are inherent in probabilistic and possibilistic methods—the former oriented toward quantifying believed or measured frequency of occurrence, and the latter seeking to determine propositions, implied by the evidence, that are similar in some sense to a hypothesis of interest. In other words, beyond accidental domain-specific relations, both types of methods are needed to analyze and clarify the significance of imprecise and uncertain information.

APPROXIMATE REASONING AND POSSIBLE WORLDS

Our point of departure is the model-theoretic formalisms of modal logics. Let us assume that declarative statements about the state, situation, or behavior

of a real-world system under study are symbolically represented by the letters of some alphabet

$$\mathscr{A} = \{p, q, r, \ldots\}$$

which are combined in the customary way using the logical operators $\neg$, $\vee$, $\wedge$, $\rightarrow$, and $\leftrightarrow$ (to be interpreted with their usual meanings) to derive a language $\mathscr{L}$ (i.e., a collection of sentences). Furthermore, we augment this language by the use of two unary operators $\mathbf{N}$ and Π, called the *necessity* and *possibility* operators, respectively, having usage governed by the rule

$$\text{If } \phi \text{ is a sentence, then } \mathbf{N}\phi \text{ and } \Pi\phi \text{ are also sentences,}$$

which introduces the ability to represent different modalities for the truth of propositions.

A *model* for this propositional system is a structure consisting of three components:

1. A nonempty set of possible worlds $\mathscr{U}$ introduced to represent states, situations, or behaviors of the system being modeled by our sentences. In what follows we will refer to this set as the *universe of discourse*, or *universe* for short.

   We will also need to consider a nonempty subset $\mathscr{E}$ of the universe $\mathscr{U}$, which is introduced to model the set of conceivable worlds that are consistent with observed evidence. This set (possibly equal to the whole universe $\mathscr{U}$) will be called the *evidential set*. Throughout this paper, we will assume that evidence about the world is always given by means of conventional propositions that allow us to determine, without ambiguity, whether a possible world either is or is not a member of the evidential set.

2. A function (called a *valuation*) that assigns one and only one of the truth values *true* or *false* to every possible world w in the universe $\mathscr{U}$ and every sentence ϕ in the language. Assignment of the truth value *true* to a pair (w, ϕ) will be denoted $w \vdash \phi$ (i.e., ϕ is true in the world w).

   In what follows, we will use the same symbols to describe subsets of possible worlds and the propositions that are true only in worlds that are members of such subsets. For example, the symbol $\mathscr{E}$ will be used to denote both the evidential set and the proposition that asserts the validity of the corresponding evidential observations. Using this notation, for example, we will write $w \vdash \mathscr{E}$ to indicate that the world w is compatible (i.e., logically consistent) with the evidence $\mathscr{E}$. Furthermore, we will use the symbol $\mathscr{L}$, introduced above as a set of well-formed sentences, to denote also the power set of the universe $\mathscr{U}$. Rigorously, subsets of $\mathscr{U}$ strictly correspond to the classes of equivalence of the sentence set $\mathscr{L}$ that are obtained by equating logically equivalent sentences. In the same simplifying vein, we will also drop the customary distinction between sentences—the linguistic expressions of something that may be true or false—and propositions—the actual things being asserted.

3. A binary relation R between possible worlds, called the *accessibility, conceivability, or reachability* relation, introduced to model the semantic of the modal operators $\mathbf{N}$ and Π.

It is not necessary to review here the well-known axioms (Hughes and Creswell [21]) that restrict the assignment of truth values to well-formed sentences according to the rules of propositional logic. To facilitate comprehension of our formalism, we need to recall solely the rules that constrain assignment of truth values to sentences formed by prefixing other valid expressions with the modal operators, that is,

1. The sentence ϕ is *necessarily true* in the possible world w (i.e., $w \vdash \mathbf{N}\phi$) if and only if it is true in every world w' that is related to the world w by the relation R.
2. The sentence ϕ is *possibly true* in the possible world w (i.e., $w \vdash \Pi\phi$) if and only if it is true in some world w' that is related to the world w by the relation R.

If, for example, the relation R relates worlds that share the same (possibly empty) subset of true sentences of the prespecified set of expressions

$$\mathcal{F} = \{\phi_1, \phi_2, \ldots\}$$

that is, if $R(w, w')$ if and only if any sentence ϕ in $\mathcal{F}$ is either true in both w and w' or false in both w and w', then the resulting system has an "epistemic" interpretation that regards related possible worlds as "being possible for all we know" (i.e., observed evidence, corresponding to a subset of $\mathcal{F}$, is the same for both worlds). In this case, the necessity operator $\mathbf{N}$ corresponds to the epistemic operator $\mathbf{K}$ of epistemic logics, with the corresponding system having the properties of the modal system S5, which was used in the context of probability theory as the semantic basis for the Dempster-Shafer [38] calculus of evidence (Ruspini [35]).

If, on the other hand, the original interpretation of logical necessity—corresponding to a relation R that is equal to $\mathcal{U} \times \mathcal{U}$, that is, that relates every pair of possible worlds—is given to the operator $\mathbf{N}$, then a proposition is necessarily true if and only if it is true in every possible world.

If the relation R is chosen as

$$R = \mathcal{E} \times \mathcal{E}$$

then this interpretation may be used to characterize approximate reasoning problems as those where a hypothesis of interest is neither necessarily true nor necessarily false in worlds in the evidential set $\mathcal{E}$, reflecting the inability of conventional deductive techniques to unambiguously determine the truth value of the hypothesis.[2]

[2] The notion of approximate reasoning problem is often extended to encompass situations where deductive techniques cannot always be used because of practical limitations on computational resources.

In those problems, in spite of this fundamental impossibility, we may resort to approximate reasoning methods to describe various properties of the evidential set $\mathscr{E}$. For example, the probabilistic structures used by various probabilistic reasoning approaches typically characterize relations of the form

$$\mu(\mathscr{H} \wedge \mathscr{E}) : \mu(-\mathscr{H} \wedge \mathscr{E})$$

between the measures of the subsets of the evidential set $\mathscr{E}$ where a hypothesis $\mathscr{H}$ is true or false, respectively.

Our aim will be to study how other structures, defining a *metric* or *distance* in the universe $\mathscr{U}$, can be used to describe the nature of the evidential set. To do so, we will assign a different meaning to the accessibility relation, giving it an interpretation that regards related worlds as "similar" or "close" in some sense. We will require, however, a scheme that is richer than that provided by a single relation so that we can extend modal notions and derive semantics bases for fuzzy logic, which relies on concepts of degrees of matching or closeness expressed by real numbers between 0 and 1.

In what follows we will use the symbols $\Rightarrow$ and $\Leftrightarrow$ to denote strong implication and equivalence, respectively. A proposition *q strongly implies p* (denoted $q \Rightarrow p$) if and only if p is true in any world where q is. Similarly, p is *logically equivalent* to q (denoted $p \Leftrightarrow q$) if and only if p and q are true in the same subset of worlds of $\mathscr{U}$.

Following traditional terminology, we will say also that a proposition p is *satisfiable* if there exists a possible world p such that $w \vdash p$.

EXTENDED MODALITIES

We first turn our attention to the problem of generalizing modal logic formalisms to explain the structures and functions of fuzzy logic.

A number of authors have studied various relations between fuzzy and modal logics. Lakoff [24], Murai et al. [28], and Schocht [36] have proposed graded generalizations of basic modal constructs. Dubois and Prade [13, 14] have also explored analogies between these nonstandard logics. In a recent paper [12], they developed, in addition, a modal basis for possibility theory by introducing fuzzy structures into modal frameworks with the goal of deriving proof mechanisms that can be used in possibilistic reasoning.

The goal for the model presented in this paper is somewhat different from the objectives guiding those efforts. We will seek explanations for possibilistic constructs on the basis of previously existing notions rather than generalizations of modal frameworks by means of fuzzy constructs. The model presented here is not based on the use of graded notions of possibility and necessity as primitive—and, by implication, easy to understand—structures. The foundation for this model is provided by a generalization of the accessibility relation,

which is given a simple interpretation as a measure of resemblance and proximity between possible worlds.

We will extend the notion of accessibility relation to encompass a family of nonempty binary relations R_α that are indexed by a numerical parameter α between 0 and 1. These relations, which are nested,

$$R_\alpha \subseteq R_\beta \qquad \text{whenever } \beta \leq \alpha$$

are introduced to represent different degrees of similarity, using a scheme that is akin to that used by Lewis in his study of counterfactuals [25]. The family of accessibility relations introduced here differs from that proposed by Lewis, however, in its use of numerical indexes[3] and in the nature of the overall modeling goals that, in Lewis's formalism, are intended to represent changes of scale induced by consideration of different restrictive statements.

Similarity Relations

To facilitate the definition of a family of accessibility relations, we introduce a *similarity function*

$$S : \mathcal{U} \times \mathcal{U} \to [0, 1]$$

assigning to each pair of possible worlds (w, w') a unique *degree of similarity* between 0 (corresponding to maximum dissimilarity) and 1 (corresponding to maximum similarity).

With the help of this function, we will then say that w and w' are related to the degree α, denoted $R_\alpha(w, w')$, if and only if $S(w, w') \geq \alpha$. In this way, the relations R_α have the required nesting property with R_0 corresponding to the whole Cartesian product $\mathcal{U} \times \mathcal{U}$ (or, every possible world is at least similar in a degree zero to every other possible world).

Some properties are required to assure that the function S has the required semantics of a metric relationship capturing the intuitive notion of similarity or "proximity." It is first necessary to demand that the degree of similarity between any world and itself be as high as possible, that is,

$$S(w, w) = 1 \qquad \text{for all } w \text{ in } \mathcal{U}$$

This property assures that every one of the accessibility relations R_α will be *reflexive* and, following the nomenclature introduced by Zadeh for fuzzy relations [52], we will also say that the similarity relation is reflexive.

Next, we will call for the function S to be *symmetric*, that is,

$$S(w, w') = S(w', w) \qquad \text{for any worlds } w \text{ and } w' \text{ in } \mathcal{U}$$

[3] We will later see that similarities can be measured by using more general, nonnumeric, scales. For simplicity reasons, I will avoid at this point the introduction of more general schemes that unnecessarily complicate the exposition.

This is a very natural requirement of any relation intended to represent a relation of resemblance between objects.

Finally, and most important, we will impose a form of *transitivity* requirement upon the similarity function S that turns it into a generalized equivalence relation. The purpose of this restriction is to assure that S has reasonable behavior as a metric in the universe of possible worlds. It would certainly be surprising if, for some similarity S, we were to be told that w and w' are very similar and that w' and w'' are also very similar, but that w does not resemble w'' at all. Clearly, there should be a lower bound on the possible values of $S(w, w'')$ that can be expressed as a function of the values of $S(w, w')$ and $S(w', w'')$. We will express such a constraint using a numeric operation, denoted $\circledast$, that takes as arguments two real numbers between 0 and 1 and returns another number in the same range, that is,

$$\circledast : [0, 1] \times [0, 1] \to [0, 1]$$

in the form of the inequality

$$S(w, w'') \geq S(w, w') \circledast S(w', w'')$$

assumed valid for any worlds w, w', and w'' in the universe $\mathcal{U}$. Reverting to a modal terminology, the above transitivity constraint, which will be called $\circledast$*-transitivity*, may be rewritten in relational form as

$$R_{\alpha \circledast \beta} \subseteq R_\alpha \circ R_\beta \qquad \text{for all } 0 \leq \alpha, \beta \leq 1$$

making obvious its generalization of the conventional definition of transitivity for ordinary binary relations, that is,

$$R \subseteq R \circ R$$

Since the role of $\circledast$, through recursive application, is that of providing a lower bound for the similarity between the two end members w_1 and w_n of a chain of possible worlds $[w_1, w_2, \ldots, w_n]$, it is obvious that the operation $\circledast$ should be commutative and associative. Furthermore, it should also be nondecreasing in each argument, as it is reasonable to ask that the desired lower bound be a monotonic function of its arguments. Finally, it is also desirable to ask that

$$\alpha \circledast 1 = 1 \circledast \alpha = \alpha$$

that is, that the values of the similarities of two indistinguishable objects to a third should be the same. These requirements are equivalent to demanding that the operation $\circledast$ be a *triangular norm* (Schweizer and Sklar [37]), or *T-norm*, for short.

Triangular norms, originally introduced in the theory of probabilistic metric spaces to treat certain statistical problems, play a distinguished role in [0, 1]-

multivalued logics (Alsina and Trillas [1], Dubois and Prade [11], Gaines [17], Rescher [31]) as the result of imposing reasonable requirements upon operations that produce the truth value of the conjunction of two expressions as a function of the truth values of the conjuncts. Furthermore, generalized similarity relations (called B-R relations by Zadeh [54]) also have an important function, to be examined further later in this paper, in the generalization of the inferential rule of modus ponens (Dubois and Prade [10], Trillas and Valverde [43]). Our axiomatic derivation for the requirement that $\circledast$ be a T-norm is based, however, solely on metric considerations, applied here to a space of possible worlds but valid in general metric spaces.

From the axioms of triangular norms, it is easy to see that

$$\alpha \circledast \beta \leq \min(\alpha, \beta)$$

which shows that the minimum function, itself a T-norm, is the largest element in this class of operations. Its minimal element, on the other hand, is the noncontinuous function $\circledast$ defined by

$$\alpha \circledast \beta = \begin{cases} \alpha & \text{if } \beta = 1 \\ \beta & \text{if } \alpha = 1 \\ 0 & \text{otherwise} \end{cases}$$

In what follows, we will also impose a most reasonable additional assumption of continuity of $\circledast$ with respect to its arguments (i.e., why should there be a jump in the value of a lower bound provided by $\circledast$ when the values of its arguments are slightly changed?). The class of continuous T-norms does not have a minimal element, although under certain additional assumptions (requiring T-norms to be also J-copulas [37]), the inequality

$$\max(\alpha + \beta - 1, 0) \leq \alpha \circledast \beta$$

also holds true, showing that certain important continuous T-norms lie between that of the $\aleph_1$-logic of Lukasiewicz (see [17]) and that of the original fuzzy logic proposed by Zadeh [53].

Continuous triangular norms play a significant part in the theories of pattern recognition and automatic classification (Ruspini [32]). In [33] I proposed the use of generalized similarity relations based on the T-norm of Lukasiewicz to generalize existing classification techniques—based on the mapping of a similarity function into a conventional equivalence relation—to the fuzzy domain by mapping these T-norms (which I called likeness relations) into generalized fuzzy partitions. Bezdek and Harris [3] independently studied axiomatic approaches to cluster analysis based on the use of continuous T-norms.

I have also studied [34] the possible relation between the multivalued logic and similarity related aspects of T-norms, and suggested that the degrees of similarity between two objects A and B may be regarded as the "degree of

truth'' of the vague proposition

''A is similar to B.''

Having argued that S should have the structure of a generalized equivalence relation, we will assume, mainly for reasons of simplicity, that the function S is the dual of a ''true'' distance, that is, that

$$S(w, w') = 1 \text{ if and only if } w = w'$$

This restriction, which is not substantial, is introduced primarily to assure that different possible worlds may be distinguished by means of the function S. Otherwise, the equivalence relation that relates two worlds w and w' if and only if $S(w, w') = 1$ may be used to partition our universe $\mathscr{U}$ into ''indistinguishable'' nonintersecting classes, indicating that our metric cannot discriminate between significant differences in system state.

Before closing our presentation of generalized similarity relations, it is important to remark upon the close relation between the notion of similarity and that of distance. If a function δ is defined in terms of a similarity function S by the simple relation

$$\delta = 1 - S$$

then it is easy to see that the function δ has the properties of a metric or distance. This is evident if the operation $\circledast$ corresponds to the T-norm of Lukasiewicz, since the transitivity condition is equivalent to the well-known triangular inequality, that is,

$$\delta(w, w'') \leq \delta(w, w') + \delta(w', w'')$$

If other T-norms are used, even stronger inequalities hold, with the so-called ''ultrametric inequality''

$$\delta(w, w'') \leq \max[\delta(w, w'), \delta(w', w'')]$$

being valid for the T-norm of Zadeh. In this case, each of the relations in the family R_α (known in fuzzy set theory as the α-cut[4] of the similarity S) is a conventional equivalence relation. This fact was exploited, prior to the introduction of fuzzy set theory and fuzzy cluster analysis, by a variety of clustering procedures of the ''single-link'' type (Jardine and Sibson [22], Sokal and Sneath [40]).

Possible and Necessary Similarity

Our semantic formalization needs require the introduction of constructs to indicate the extent by which a concept exemplifies, illustrates, or is an

[4]The α-cut [46] of a fuzzy set $\mu: \mathscr{U} \to [0, 1]$ is the conventional set of all points w such that $\mu(w) \geq \alpha$. A similar concept is defined for relations as subsets of a product space $\mathscr{U} \times \mathscr{U}$.

adequate model of another concept. Our interpretations will therefore be oriented toward characterization of the degree to which a concept can be said to be a good example of another concept with the purpose of defining vague concepts by means of measures of proximity between defined and defining concepts. In our treatment, each of the multiple "definiens" will be a conventional proposition corresponding to a subset of possible worlds. It is conceivable, however, that new vague concepts might also be described metric relations to other vague concepts.

The required constructs are based on the idea that whenever p and q are propositions such that $p \Rightarrow q$, then any p-world is an "example" of a q-world. This basic notion will be generalized by the introduction of modal structures that define to what degree possible worlds that satisfy a certain proposition q fit a vague concept. Some of those possible worlds are "paradigmatic" of the vague concept, that is, they fit it to a degree equal to 1 in the same sense that we may say, for example, that somebody whose height is 7 ft is definitely "tall." If we use a notion of graded fitness, however, certain worlds will fit the concept to a degree, that is, they resemble (or are similar to) some paradigmatic example of the vague concept.

The conventional interpretation of possibility must be modified, therefore, to capture the idea that a particular possible world is similar in some degree to another world that satisfies a "reference" proposition.

More generally, however, we will be interested in relations of similarity between *pairs of subsets* of possible worlds rather than between pairs of possible worlds. This requirement complicates matters considerably, because we will be forced to consider both the "validity" of a proposition p in *some* world where another proposition q is true and its applicability in *every* world where q is true. In the former case, we will care about the existence of q-worlds that are similar to some degree to some p-world, whereas in the latter we will be concerned with the size of the minimum neighborhood of p (as a subset of the universe $\mathscr{U}$) that fully encloses the subset q.

This dual concern for what may possibly apply and what must necessarily hold—an essential aspect of modal logic—is typical of situations where relationships between ensembles of objects are described in terms of relations between their members. In the probability calculus, for example, knowledge of probabilities over certain families of subsets provides "sharp" upper and lower bounds (called *lower* and *upper probabilities*, respectively) for the probabilities of other subsets—an important fact in the extension of set measures to larger domains (Halmos [19]). The role and properties of these bounds in the Dempster-Shafer [38] calculus of evidence is well known, having been described in the original paper of Dempster [8], related to concepts of modal logic by Ruspini [35], and being also the subjects of considerable formal study (Choquet [7]) as mathematical structures.

Analogies between the role of probabilistic bounds (i.e., bounds for probability values) and possibility/necessity distributions, have been the source of much of the confusion about the need for possibilistic schemes. Each upper/lower-bound pair, however, leads to a substantially different description of the nature of a subset of possible worlds, being, in either case, measures that arise naturally when pointwise properties are extended to set partitions. General properties of these measures have been studied by Dubois and Prade [15] in the context of approximate reasoning and in other regards by Pavlak [30].

Our generalizations of the notions of possibility and necessity are related to the so-called *de re* (Hughes and Creswel [21]) interpretation of the statement "If q, then p is possible" as the modal propositional relation

$$q \Rightarrow \Pi p$$

We will say that the proposition q implies, or is a *necessary model* of, the proposition p to the degree α if and only if for every q-world w there exists a p-world w' that is at least α-similar to it [i.e., $S(w, w') \geq \alpha$] or, equivalently, whenever

$$q \Rightarrow \Pi_\alpha p$$

Similarly, we will say that the proposition q is *consistent with*, or is a *possible model* of, the proposition *p to the degree* α[5] if and only there exist a q-world w and a p-world w' that are at least α-similar or, equivalently, whenever

$$\neg(p \Rightarrow \neg\Pi_\alpha q)$$

The similarity function that we have introduced in the universe $\mathscr{U}$ provides us with a simple mechanism to quantify both the extent of "inclusion" and that of the "intersection" between pairs of subsets of possible worlds.[6]

Possibilistic Implication and Consistency

The notion of subset inclusion and its related concept of set identity are of central importance in deductive logic, since subsets of possible worlds are formally equivalent to propositions with subset inclusion and identity corre-

[5]Note that our characterizations of both possibility and necessity distributions are based in the modal possibility operators Π_α.

[6]For reasons that by now should be evident, we will not need to introduce a concept of "unconditioned possibility" although it would be easy to do so using $q = \mathscr{U}$. Being concerned with the power of certain propositions to exemplify other conditions, we will not have much occassion to deal with the strength of tautologies in that regard.

sponding to logical implication and equivalence, respectively. These propositional relationships are the basis of derivation rules such as the modus ponens. The notion of intersection plays a similar role in modal analyses because of its ability to express the potential validity of a statement.

Classical accounts, however, recognize only two "degrees" of inclusion corresponding to the cases when either a set q is a subset of another set p or it is not, with a similar dichotomy applying to degrees of intersection. Our generalization exploits the metric structures defined between sets of possible worlds by introducing measures that describe a subset as enclosed in a *neighborhood* (of some size) of another set while intersecting another of its neighborhoods (of "smaller" size).[7] The problem of measuring the "size" of those neighborhoods is the subject of our immediate considerations.

DEGREE OF IMPLICATION Our definition of partial implication between propositions was based on conditions that determine whether, given two propositions p and q, one of them implies the other to the same value α. In particular, since every world w is always similar in a degree that is at least equal to zero to any other world w', it is always true that any proposition q implies any other proposition p to the degree zero. It is often the case, however, that the degree of implication between p and q is at least equal to some certain positive value α.

If we want to generalize procedures based on inclusion relationships, such as the modus ponens, in an efficient fashion, we will need to measure the "optimal" (or maximum) value of the parameter α such that q implies p to the degree α. This value is a measure of the degree to which the set of all p-worlds must be "stretched" to encompass the set of all q-worlds. The least upper bound of the values of the similarities between any q-world w' and some p-world w is given by the *degree of implication* function:

DEFINITION 1 *The* degree of implication *of p by q is the value*

$$\mathrm{I}(p \mid q) = \inf_{w' \vdash q} \sup_{w \vdash p} S(w, w')$$

Defined in this way, the degree of implication $\mathrm{I}(p \mid q)$ is a measure of the "minimal amount" of stretching required to reach a p-world from any q-world, in the sense that if $\beta < \mathrm{I}(p \mid q)$, then

$$q \Rightarrow \Pi_\beta p$$

[7] It is important to recall that, owing to our reliance on similarity rather than on the dual notion of dissimilarity or distance, high values of α correspond to low values of "stretching" or to smaller set neighborhoods.

Furthermore, α is the largest real value for which the above statement may be made.

As the following theorem makes clearer, this function provides the basis for the generalization of the modus ponens. This truth-derivation procedure may be thought of as an expression of the nesting relationships that hold between the sizes of neighborhoods of such subsets.

THEOREM 1 *The degree of implication function,*

$$\mathrm{I}: \mathscr{L} \times \mathscr{L} \to [0, 1]$$

has the following properties:

(i) *If $p \Rightarrow r$, then $I(p \mid q) \leq I(r \mid q)$*

(ii) *If $p \Rightarrow r$, then $I(p \mid q) \geq I(p \mid r)$*

(iii) *$I(p \mid q) \geq I(p \mid r) \circledast I(r \mid q)$*

where p, q, and r are any satisfiable propositions.

Proof The first two properties are an immediate consequence of the definition of degree of implication. To prove the third, observe that by definition of similarity,

$$S(w, w') \geq S(w, w'') \circledast S(w'', w')$$

for any worlds w, w', and w''.

Taking the supremum on both sides of this inequality with respect to all worlds $w \vdash p$, it follows, because $\circledast$ is continuous, that

$$\sup_{w \vdash p} S(w, w') \geq \left[\sup_{w \vdash p} S(w, w'')\right] \circledast S(w'', w')$$

Since this expression is true, in particular, for all worlds $w'' \vdash r$, it is true that

$$\sup_{w \vdash p} S(w, w') \geq \left[\inf_{w'' \vdash r} \sup_{w \vdash p} S(w, w'')\right] \circledast S(\hat{w}, w')$$

$$= \mathrm{I}(p \mid r) \circledast S(\hat{w}, w')$$

where $\hat{w}$ is any world such that $\hat{w} \vdash r$.

From this inequality, it follows, since $\circledast$ is continuous, that

$$\sup_{w \vdash p} S(w, w') \geq \mathrm{I}(p \mid r) \circledast \left[\sup_{\hat{w} \vdash r} S(\hat{w}, w')\right]$$

Taking now the infimum on both sides of this expression over all worlds w' such that $w' \vdash q$, it is easy to see, using again the continuity of $\circledast$, that

$$\inf_{w' \vdash q} \sup_{w \vdash p} S(w, w') \geq \mathrm{I}(p \mid r) \circledast \left[\inf_{w' \vdash q} \sup_{\hat{w} \vdash r} S(\hat{w}, w')\right]$$

proving the $\circledast$-transitivity of I. ■

Note, that since $\mathrm{I}(q \mid q) = 1$ for any proposition q, the following statement is also true.

COROLLARY *If p and q are propositions in $\mathscr{L}$, then*

$$\mathrm{I}(p \mid q) = \sup_r \left[\mathrm{I}(p \mid r) \circledcirc \mathrm{I}(r \mid q)\right]$$

Notice also that if $\mathrm{I}(p \mid q) = 1$, then

$$\sup_{w \vdash p} S(w, w') = 1 \qquad \text{for all } w' \vdash q$$

Under minimal assumptions (assuring that the supremum operation is actually a maximization), this relation is equivalent to stating that q strongly implies p, or that any q-world is also a p-world.

The nonsymmetric function I measures the extent to which every world w' in a certain class resembles some world w (dependent on w') in a reference class, explicating the nature of the nonsymmetric assessments (Tversky [44]) found in psychological experimentation when subjects are asked to evaluate the degree to which an object "resembles" another. The results obtained in those experiments suggest that human beings, when assessing similarity between objects, use one of them (or a class of similar objects) as a reference landmark to describe the other. Such asymmetries might be explained by noticing that, in general, $\mathrm{I}(p \mid q) \neq \mathrm{I}(q \mid p)$, indicating that the stronger stimulus might generally be used to construct a reference class, which is then used to describe other stimuli.

The degree of implication of one proposition by another can be readily used to generate a measure of similarity between propositions that generalizes our original measure of similarity between possible worlds:

$$\hat{S}(p, q) = \min\left[\mathrm{I}(p \mid q), \mathrm{I}(q \mid p)\right]$$

quantifying the degree by which the propositions p and q are equivalent. It can be readily proved (Valverde [45]) from its definition and from the transitivity property of I that $\hat{S}$ is a reflexive, symmetric, and $\circledcirc$-transitive function between subsets of possible worlds. This similarity function is the dual of the well-known *Hausdorff distance*, defined between subsets of a metric as a function of the distance between pairs of their members (Dieudonné [9]), which is given by the expression

$$\hat{\delta}(A, B) = \max\left\{\left[\sup_{x \in A} \inf_{y \in B} \delta(x, y)\right], \left[\sup_{x \in B} \inf_{y \in A} \delta(x, y)\right]\right\}$$

The result expressed by the transitive property of the degree of implication may be stated using modal notation in the form

$$q \Rightarrow \Pi_\alpha r \text{ and } r \Rightarrow \Pi_\beta p \quad \text{imply that} \quad q \Rightarrow \Pi_{\alpha \circledcirc \beta} p$$

as the simplest form of the generalized modus ponens rule of Zadeh.

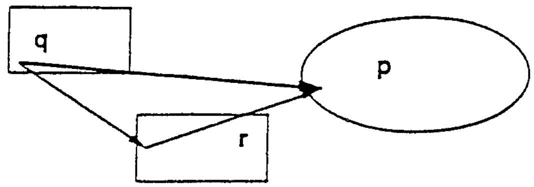


Figure 1. The generalized modus ponens.

The relationship between this rule and the classical modus ponens is easier to perceive if it is remembered that classical conditional propositions of the form "If q, then p" simply state that the set of q-worlds is a subset of the set of p-worlds. Such relationships of inclusion can also be described in metric terms by saying that every q-world has a p-world (i.e., itself) that is as similar as possible to it.

Logic structures, however, only allow us to say either that q implies p or that q implies its negation $\neg p$, or that neither of those statements is true. By contrast, similarity relations allow measurement of the amount by which a set must be "stretched" (as illustrated in Figure 1) to enclose another set. Using such metrics, we can describe the generalized modus ponens as a relation between the stretching required to reach p from any point of the set r, the stretching required to reach r from any point of the set q, and the stretching required to reach p from any point of the set q.

In the section Generalized Inference, we will derive alternative expressions for the generalized modus ponens that allow us to propagate both measures characterizing degree of implication and degree of consistency; a dual concept that plays, with respect to the notion of possibility, the function that is fulfilled by the degree of implication function with respect to necessity. In those derivations, by introducing sharper bounds for certain conditional concepts, we will also be able to improve the quality of the bounds provided by generalized modus ponens rules while being closer in spirit to its usual fuzzy-logic formulation.

DEGREE OF CONSISTENCY A notion that is dual to that of degree of implication is given by a function that measures the pointwise proximity between pairs of possible worlds from an "optimistic" point of view characterizing the degree to which statements that are true in some worlds *may* apply in others. By contrast, the degree of implication measures the extent to which statements that are true in p-worlds *must* hold in q-worlds.

DEFINITION 2 *The* degree of consistency *of p and q is the value*

$$C(p \mid q) = \sup_{w' \vdash q} \sup_{w \vdash p} S(w, w')$$

An immediate consequence of this definition that $C(\cdot \mid \cdot)$ is a symmetric

function that is increasingly monotonic in both arguments (with respect to the $\Rightarrow$). It is also easy to see that the values of the degree of consistency function are never smaller than the corresponding values of the degree of consistence function,

$$\mathbf{I}(p \mid q) \leq \mathbf{C}(p \mid q)$$

as the amount of stretching required to reach p from some "convenient" q-world is smaller (i.e., higher values of S) than that required to reach p from any q-world. In general, however, the degree of consistency function is not transitive, preventing the statement of a "compatibility" counterpart of the generalized modus ponens rule. Its relationship with the degree of implication function expressed by the expression

$$\mathbf{C}(p \mid q) = \sup_{w' \vdash q} \mathbf{I}(p \mid w') = \sup_{w \vdash p} \mathbf{I}(q \mid w)$$

will permit us, nonetheless, to derive a useful bound-propagation expression.

POSSIBILITY AND NECESSITY DISTRIBUTIONS

This section presents interpretations of the major constructs of fuzzy logic—possibility and necessity distributions—in terms of similarity-based structures. Possibility and necessity distributions are functions that measure the proximity of either all or some of the worlds in the evidential set to worlds in other sets that are employed as reference landmarks.

The role played by possibility and necessity distributions is similar to that performed by lower and upper bounds of probability distributions (or by the belief and plausibility functions of the Dempster–Shafer calculus of evidence) with respect to probability distributions. The essential difference between these bounds and those provided by possibility/necessity pairs lies in the fundamentally dissimilar character of what is being bound—metric structures relating pairs of worlds in one case and measures of set size in the other. Furthermore, in the model of possibilistic structures that is presented in this paper, necessity (possibility) distributions are any lower (upper) bounds of a certain metric function rather than its "best" or "sharpest" bounds. The operations of fuzzy logic allow computation of bounds for some of these measures as a function of bounds of other measures.

Inverse of a Triangular Norm

When working in ordinary metric spaces, it is often convenient to express the conventional statement of the triangular inequality,

$$\delta(w, w') \leq \delta(w, w'') + \delta(w'', w')$$

Table 1. Triangular Norms, Conorms, and Pseudoinverses

Name	T-Norm $a \circledast b$	Conorm $a \oplus b$	Pseudoinverse $a \oslash b$
Lukasiewicz	$\max(a + b - 1, 0))$	$\min(a + b, 1)$	$\min(1 + a - b, 1)$
Product	ab	$a + b - ab$	a/b if $b > a$ 1 otherwise
Zadeh	$\min(a, b)$	$\max(a, b)$	a if $b > a$ 1 otherwise

in the equivalent form

$$\delta(w, w') \geq |\delta(w, w'') - \delta(w', w'')|,$$

which utilizes a form of inverse (i.e., the subtraction operator $-$) of the function used to express the original inequality (i.e., the addition operator $+$). This notion of inverse can be directly generalized (Schweizer and Sklar [37]) to provide us with the tools required to define possibility and necessity functions and to derive useful forms of the generalized modus ponens involving either type of these constructs.

Definition 3 *If* $\circledast$ *is a triangular norm, its* pseudoinverse $\oslash$ *is the function defined over pairs of numbers in the unit interval of the real line by the expression*

$$a \oslash b = \sup\{c : b \circledast c \leq a\}$$

From this definition it is clear that $a \oslash b$ is nondecreasing in a and nonincreasing in b. Furthermore, $a \oslash 0 = 1$ and $a \oslash 1 = a$ for any a in $[0, 1]$. Other important properties of the pseudoinverse function are given in the works of Schweizer and Sklar [37], Trillas and Valverde [43], and Valverde [45].

Examples of the pseudoinverses of important triangular norms are given in Table 1 together with the corresponding conorms.

Unconditioned Necessity Distributions

We introduce first a family of functions that bound from below the value of the similarity between any evidential world in $\mathscr{E}$ and some world where another proposition p is true. These *unconditioned necessity* distributions are lower bounds for values of the degree of implication $\mathbf{I}(p \mid \mathscr{E})$, which measures the extent to which statements that are true in a reference set (i.e., the subset of p-worlds) must hold in the evidential set.

As observed before, whenever $\mathbf{I}(p \mid \mathscr{E}) = 1$, it is true, under minimal assumptions, that the evidential subset $\mathscr{E}$ is a subset of the set of all p-worlds,

or that p necessarily holds in $\mathscr{E}$. If, on the other hand, $\mathbf{I}(p \mid \mathscr{E}) = \alpha < 1$, then p must be stretched a certain amount—with smaller α corresponding to greater stretching—in order for one of its neighborhoods to encompass $\mathscr{E}$.

Definition 4 *If $\mathscr{E}$ is an evidential set, then a function* Nec($\cdot$) *defined over propositions in the language $\mathscr{L}$ is called an* unconditioned necessity distribution *for $\mathscr{E}$ if*

$$\mathrm{Nec}(p) \leq \mathbf{I}(p \mid \mathscr{E})$$

Unconditioned Possibility Distributions

The dual counterpart of the unconditioned necessity distribution is provided by upper bounds of the degree of consistency $\mathbf{C}(p \mid \mathscr{E})$. Whenever $\mathbf{C}(p \mid \mathscr{E}) = 1$, it is easy to see that, under minimal assumptions, there exists a p-world w that is in the evidential set $\mathscr{E}$ or, equivalently, that p (for all we know) is possibly true. If, on the other hand, $\mathbf{C}(p \mid \mathscr{E}) = \alpha < 1$, then there exists a neighborhood (of "size" α) of some p-world that intersects the evidential set.

Definition 5 *If $\mathscr{E}$ is an evidential set, then a function* Poss($\cdot$) *defined over propositions in the language $\mathscr{L}$ is called an* unconditioned possibility distribution *for $\mathscr{E}$ if*

$$\mathrm{Poss}(p) \geq \mathbf{C}(p \mid \mathscr{E})$$

Since the value Poss(p) of any possibility function Poss($\cdot$) is an upper bound of the value $\mathrm{C}(p \mid \mathscr{E})$ of the degree of consistence, the corresponding value Nec(p) of any necessity function Nec($\cdot$) is a lower bound of $\mathrm{I}(p \mid q)$, it follows that values of a possibility function can never be smaller than the corresponding values of any necessity function, that is, that

$$\mathrm{Nec}(p) \leq \mathrm{Poss}(p)$$

Properties of Possibility and Necessity Distributions

In this subsection we will develop similarity-based interpretations for some basic formulas of possibilistic calculus. These expressions may be thought of as mechanisms that allow the extension of a partially known possibility distribution. For example, the property that

$$\max[\mathrm{Poss}(p), \mathrm{Poss}(q)] \geq \mathbf{C}(p \vee q \mid \mathscr{E})$$

which is proved below, is the similarity interpretation of the standard rule that allows computation of the value of the possibility of a disjunction in fuzzy logic, that is,

$$\mathrm{Poss}(p \vee q) = \max[\mathrm{Poss}(p), \mathrm{Poss}(q)]$$

Theorem 2 *If p and q are propositions, and if the quantities $Poss(p)$, $Poss(q)$, $Nec(p)$, and $Nec(q)$ are such that*

$$\mathrm{Nec}(p) \le \mathrm{I}(p \mid \mathcal{E}), \qquad \mathrm{Nec}(q) \le \mathrm{I}(q \mid \mathcal{E})$$

$$\mathrm{Poss}(p) \ge \mathrm{C}(p \mid \mathcal{E}), \qquad \mathrm{Poss}(q) \ge \mathrm{C}(q \mid \mathcal{E})$$

then the following statements (similarity-based interpretations of the basic laws of fuzzy logic) are valid:

$$\max[\mathrm{Nec}(p), \mathrm{Nec}(q)] \le \mathrm{I}(p \vee q \mid \mathcal{E})$$

$$\max[\mathrm{Poss}(p), \mathrm{Poss}(q)] \ge \mathrm{C}(p \vee q \mid \mathcal{E})$$

$$\min[\mathrm{Poss}(p), \mathrm{Poss}(q)] \ge \mathrm{C}(p \wedge q \mid \mathcal{E})$$

Proof Note first that since $\mathrm{C}(\cdot \mid \cdot)$ is nondecreasing (with respect to the $\Rightarrow$ order) in its arguments, it is true that

$$\mathrm{Poss}(p) \ge \mathrm{C}(p \mid \mathcal{E}) \ge \mathrm{C}(p \wedge q \mid \mathcal{E})$$

$$\mathrm{Poss}(q) \ge \mathrm{C}(q \mid \mathcal{E}) \ge \mathrm{C}(p \wedge q \mid \mathcal{E})$$

whenever $p \wedge q$ is satisfiable, from which it is easy to see that

$$\min[\mathrm{Poss}(p), \mathrm{Poss}(q)] \ge \mathrm{C}(p \wedge q \mid \mathcal{E})$$

The corresponding result is obvious when $p \wedge q$ is nonsatisfiable.

A similar argument shows, for necessity functions, that

$$\max[\mathrm{Nec}(p), \mathrm{Nec}(q)] \le \mathrm{I}(p \vee q \mid \mathcal{E})$$

To prove the disjunctive law for possibilities, notice that if f is any function mapping elements of a general domain D into real numbers, then

$$\sup\{f(d) : d \in A \cup B\} = \max[\sup\{f(d) : d \in A\}, \sup\{f(d) : d \in B\}]$$

From this equality, it is easy to see that if $\mathrm{Poss}(p)$ and $\mathrm{Poss}(q)$ are upper bounds of $\mathrm{I}(p \mid \mathcal{E})$ and $\mathrm{I}(q \mid \mathcal{E})$, respectively, then

$$\max[\mathrm{Poss}(p), \mathrm{Poss}(q)] \ge \mathrm{C}(p \vee q \mid \mathcal{E})$$

which completes the proof of the theorem. ■

Note, however, that another law commonly given as an axiom for necessity functions does not hold valid in our interpretation. As illustrated in Figure 2, the distance from a point to the intersection of two sets may be strictly larger than the distance to either set (i.e., the similarity will be strictly smaller). In general, therefore,

$$\min[\mathrm{Nec}(p), \mathrm{Nec}(q)] \not\le \mathrm{I}(p \wedge q \mid \mathcal{E})$$

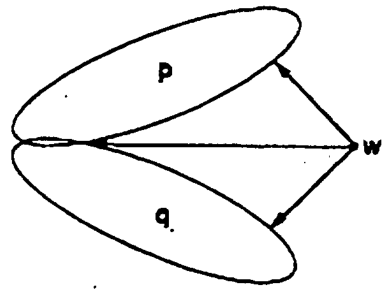


Figure 2. Failure of conjunctive necessity.

making invalid, under this interpretation, the conjunctive law for necessities (Dubois and Prade [11])

$$\mathrm{Nec}(p \wedge q) = \min[\mathrm{Nec}(p), \mathrm{Nec}(q)]$$

We may also note in this regard that the similarity-based model that is discussed here does not make use of the notion of negation either as a mechanism to generate dual concepts or in its own right as an important logical concept. It is my intent to study, in the immediate future, alternative models in which notions of negation and maximal dissimilarity play more substantive roles.

Conditional Possibilities and Necessities

The concepts of conditional possibility and necessity are closely related to the previously introduced unconditioned structures. These structures may be thought of as a characterization of the proximity of a world w to some or all of the worlds where a proposition p is true, *given that* w *is similar in the degree 1 to the evidential set* $\mathscr{E}$ (i.e., $w \vdash \mathscr{E}$). With this fact in mind, we could have used the somewhat baroque formulation

$$\mathrm{C}(p \mid \mathscr{E}) = \sup_{w \vdash \mathscr{E}} [\mathrm{I}(p \mid w) \oslash \mathrm{I}(\mathscr{E} \mid w)]$$

to define unconditioned possibility distributions—a rather unnecessary effort if we consider that $\mathrm{I}(\mathscr{E} \mid w) = 1$ whenever $w \vdash \mathscr{E}$, showing its obvious equivalence to the simpler form used in the previous section. In spite of such observation, the above identity is important in understanding the purpose of the definitions that follow. Those definitions interpret conditional possibilities and necessities as a measure of the proximity of worlds on the evidential set $\mathscr{E}$ to (some or all) worlds satisfying a (conditioned) proposition p relative to their proximity to (some or all of) the worlds that satisfy another (conditioning) proposition q.

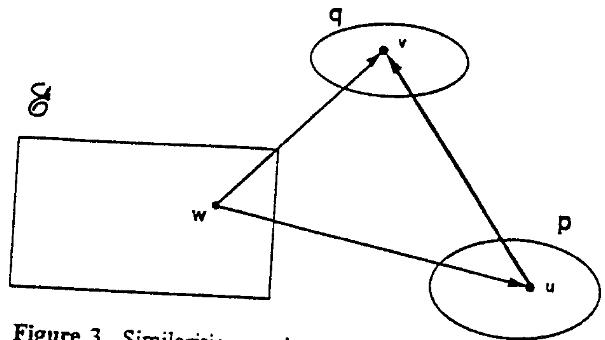


Figure 3. Similarities as viewed from the evidential set.

The mechanism used to specify that relationship, which is closely related in spirit to results of Valverde [45] on the structure of indistinguishability relations, is based on the pseudoinverse function introduced earlier. The basic idea used by these definitions is also illustrated in Figure 3, where, from the perspective of the evidential world w, the similarity between the p-world u and the q-world v is estimated by means of an inequality that generalizes the "absolute value" form of the triangular inequality,

$$\delta(u, v) \geq |\,\delta(u, w) - \delta(v, w)\,|$$

to its similarity-based form

$$S(u, v) \leq \min[S(u, w) \oslash S(v, w), S(v, w) \oslash S(u, w)]$$

The required interplay between similarities to conditioning and conditioned sets is captured by the following definitions.

DEFINITION 6 *Let $\mathscr{E}$ be an evidential set. A function $Nec(\cdot \mid \cdot)$ mapping pairs of propositions in the language $\mathscr{L}$ into* [0, 1] *is called a* conditional necessity distribution *for $\mathscr{E}$ if*

$$\mathrm{Nec}(q \mid p) \leq \inf_{w \vdash \mathscr{E}} [\mathrm{I}(q \mid w) \oslash \mathrm{I}(p \mid w)]$$

for any propositions p and q in $\mathscr{L}$.

DEFINITION 7 *Let $\mathscr{E}$ be an evidential set. A function $Poss(\cdot \mid \cdot)$ mapping pairs of propositions in the language $\mathscr{L}$ into* [0, 1] *is called a* conditional possibility distribution *for $\mathscr{E}$ if*

$$\mathrm{Poss}(q \mid p) \geq \sup_{w \vdash \mathscr{E}} [\mathrm{I}(q \mid w) \oslash \mathrm{I}(p \mid w)]$$

for any propositions p and q in $\mathscr{L}$.

It is easy to see from these definitions that the values of a conditional

necessity distribution are never larger than the corresponding values of a conditional possibility distribution, that is,

$$\mathrm{Nec}(q \mid p) \leq \mathrm{Poss}(q \mid p)$$

Furthermore, since $\mathbf{I}(\cdot \mid \cdot)$ is $\circledast$-transitive, it is

$$\mathbf{I}(q \mid w) \geq \mathbf{I}(q \mid p) \circledast \mathbf{I}(p \mid w)$$

From this inequality and the definition of psuedoinverse of a triangular norm, it is easy to see that $\mathbf{I}(q \mid p)$ is a conditional necessity function, showing also that the bounds provided by the evidential-set perspective are better than those that can be obtained by direct use of the degree of implication as the definition of conditional necessity.[8]

Note also that if $\mathrm{Nec}(p) = 1$, indicating that $\mathbf{I}(p \mid \mathscr{E}) = 1$, and if $\mathrm{Nec}(q \mid p) = 1$, then the above definition of conditional necessity shows that $\mathbf{I}(q \mid \mathscr{E}) = 1$, indicating that $\mathrm{Nec}(q)$ may be taken to be equal to 1, thus generalizing the well-known axiom (consequential closure) of certain modal systems (e.g., the system **T**, as discussed in Hughes and Creswell [21])

$$\text{If } \mathrm{N}p \text{ and } \mathrm{N}(p \to q), \text{ then } \mathrm{N}q.$$

The definitions above can also be further interpreted as a way to compare the similarities between evidential worlds and those in the conditioning and conditioned sets by noting that whenever

$$\mathbf{I}(q \mid w) \geq \mathbf{I}(p \mid w)$$

for every evidential world $w \vdash \mathscr{E}$, then $\mathrm{Nec}(q \mid p)$ may be chosen to be equal to 1. Similarly, if there exists some world $w \vdash \mathscr{E}$ where this inequality holds, then it is $\mathrm{Poss}(q \mid p) = 1$. In either case, however, the maximum value for the conditional distribution (i.e., 1) is reached when the proximity of one evidential world w, in the case of possibilities, or of every one of them, in the case of necessities, to a world w_q in the conditioned set exceeds the proximity of w to the conditioning set p. In either case, once again returning to an apparent notational overkill, we may state this fact by means of the identity function τ in the unit interval:

$$\tau : [0, 1] \mapsto [0, 1] : \alpha \mapsto \alpha$$

in the form

$$\mathbf{I}(q \mid w) \geq \tau(\mathbf{I}(p \mid w))$$

for some $w \vdash \mathscr{E}$ in the case of possibilities, with the same inequality holding

[8] A dual result for possibilities involving $\mathrm{C}(q \mid p)$ does not hold in general. It is easy to see, however, that $\mathrm{C}(q \mid \mathscr{E}) \oslash \mathbf{I}(p \mid \mathscr{E})$ is a possibility function for q given p.

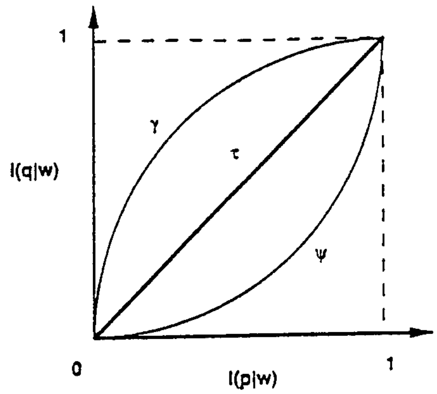


Figure 4. Examples of possible similarity relationships between conditioning and conditioned sets.

for every $w \vdash \mathscr{E}$ in the case of necessities. We can, however, conceive of other functions

$$\gamma : [0,1] \mapsto [0,1] : \alpha \mapsto \gamma(\alpha)$$

with $\gamma(\alpha) \geq \alpha$ to specify a stronger form of implication, as illustrated in Figure 4, that is,

$$\mathrm{I}(q \mid w) \geq \gamma(\mathrm{I}(p \mid w))$$

Similarly, we can also conceive of functions ψ with $\psi(\alpha) \leq \alpha$ that can be used to model weaker forms of implication.

Possibilistic calculi based on the propagation of truth mappings of this type, first proposed by Baldwin [2], are utilized in the RUM (Bonissone and Decker [4], Bonissone et al. [5]) and MILORD (Godo et al. [18]) expert systems. The particular case when $\gamma = \tau$, stating that every α-cut of the conditioning proposition p is fully enclosed (in the conventional sense) in the α-cut of the conditioned proposition q, has been called *truth mapping* in fuzzy logic literature.

The primary purpose of conditional distributions, however, is to provide a quantitative measure of the degree to which one proposition may be said to imply another with a view to extending inferential procedures by means of structures that superimpose the topological notion of continuity upon a logical framework concerned with propositional validity.

GENERALIZED INFERENCE

The major inferential tool of fuzzy logic is the *compositional rule of inference* of Zadeh [53], which generalizes the corresponding classical rule of

inference by its ability to infer valid statements even when a perfect match between facts and rule antecedent does not exist, that is, from

$$\frac{\begin{array}{c} p \\ p \rightarrow q \end{array}}{q}$$

to its "approximate" version

$$\frac{\begin{array}{c} p' \\ p \rightarrow q \end{array}}{q'}$$

where p' and q' are similar to p and q, respectively. In this sense, the generalized modus ponens operates as an "interpolation" (or, more precisely, as an "extrapolation") procedure in possible-world space.

Unlike the interpolation procedures of numerical analysis, however, which yield estimates of function value, this extrapolation procedure approximates truth in the sense that it produces a proposition that is more general than the consequent of the inferential rule but resembles it to some degree (which is a function of the degree to which p' resembles p). The "extrapolated conclusion," however, is a correctly derived proposition, that is, the result of a sound logical procedure rather than of an approximate heuristic technique.

Generalized Modus Ponens

The theorems that are proved below are based on the use of a family $\mathscr{P}$ of propositions that partitions the universe of discourse $\mathscr{U}$ in the sense that every possible world will satisfy at least one proposition in $\mathscr{P}$.

DEFINITION 8 *If $\mathscr{P}$ is a subset of satisfiable propositions in $\mathscr{L}$ such that if w is a possible world in the universe $\mathscr{U}$, then there exists a proposition p in $\mathscr{P}$ such that $w \vdash p$, then the family $\mathscr{P}$ is called a* partition *of $\mathscr{U}$.*

These results make use of information such as the values of the unconditioned necessity or possibility distributions for antecedent propositions p in the family $\mathscr{P}$ together with the values $\mathrm{Nec}(q \mid p)$ or, respectively, $\mathrm{Poss}(q \mid p)$ to "extend" the unconditioned distributions to the "consequent" proposition q. In this sense, these findings interpret, in the same spirit used in Theorem 2 for other basic laws, the generalized modus ponens laws of fuzzy logic:

$$\mathrm{Nec}(q) = \sup_{\mathscr{P}} \left[\mathrm{Nec}(q \mid p) \circledcirc \mathrm{Nec}(p)\right]$$

$$\mathrm{Poss}(q) = \sup_{\mathscr{P}} \left[\mathrm{Poss}(q \mid p) \circledcirc \mathrm{Poss}(p)\right]$$

THEOREM 3 (GENERALIZED MODUS PONENS FOR NECESSITY FUNCTIONS) *Let $\mathscr{P}$ be a partition of $\mathscr{U}$ and let q be a proposition. If $\mathrm{Nec}(p)$ and*

$Nec(q \mid p)$ are real values defined for every proposition p in the partition $\mathcal{P}$ such that

$$\mathrm{Nec}(p) \leq \mathbf{I}(p \mid \mathcal{E})$$

$$\mathrm{Nec}(q \mid p) \leq \inf_{w \vdash \mathcal{E}} \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid w)\right]$$

then the following inequality is valid:

$$\sup_{\mathcal{P}} \left[\mathrm{Nec}(q \mid p) \circledast \mathrm{Nec}(p)\right] \leq \mathbf{I}(q \mid \mathcal{E})$$

Proof Note first that since $\oslash$ is nonincreasing in its second argument and since

$$\mathbf{I}(p \mid \mathcal{E}) \leq \mathbf{I}(p \mid w)$$

for every evidential world w,

$$\mathrm{Nec}(q \mid p) \leq \inf_{w \vdash \mathcal{E}} \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid w)\right] \leq \inf_{w \vdash \mathcal{E}} \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid \mathcal{E})\right]$$

It follows then from the monotonicity and continuity of $\circledast$ with respect to its arguments that

$$\begin{aligned}
\mathrm{Nec}(p) \circledast \mathrm{Nec}(q \mid p) &\leq \mathbf{I}(p \mid \mathcal{E}) \circledast \inf_{w \vdash \mathcal{E}} \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid \mathcal{E})\right] \\
&= \inf_{w \vdash \mathcal{E}} \left\{\mathbf{I}(p \mid \mathcal{E}) \circledast \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid \mathcal{E})\right]\right\} \\
&\leq \inf_{w \vdash \mathcal{E}} \mathbf{I}(q \mid w) \\
&= \mathbf{I}(q \mid \mathcal{E})
\end{aligned}$$

since

$$\mathbf{I}(p \mid \mathcal{E}) \circledast \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid \mathcal{E})\right] \leq \mathbf{I}(q \mid w)$$

because of the definition of $\oslash$ and the continuity of $\circledast$.

Since the above inequality is valid for any proposition p in $\mathcal{P}$, Theorem 3 follows. ■

A dual result also holds for possibility functions.

THEOREM 4 (GENERALIZED MODUS PONENS FOR POSSIBILITY FUNCTIONS *Let $\mathcal{P}$ be a partition of $\mathcal{U}$ and let q be a proposition. If $Poss(p)$ and $Poss(q \mid p)$ are real values, defined for every proposition p in $\mathcal{P}$, such that*

$$\mathrm{Poss}(p) \geq \mathbf{C}(p \mid \mathcal{E})$$

$$\mathrm{Poss}(q \mid p) \geq \sup_{w \vdash \mathcal{E}} \left[\mathbf{I}(q \mid w) \oslash \mathbf{I}(p \mid w)\right]$$

then the following inequality is valid:

$$\sup_{\mathscr{P}} \left[\operatorname{Poss}(q \mid p) \circledast \operatorname{Poss}(p)\right] \geq \mathrm{C}(q \mid \mathscr{E})$$

Proof Note first that if w is an evidential world, then

$$\mathrm{C}(p \mid \mathscr{E}) \geq \mathrm{I}(p \mid w)$$

It follows then from the nonincreasing nature of $\oslash$ with respect to its second argument that

$$\begin{aligned}\operatorname{Poss}(q \mid p) &\geq \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{I}(p \mid w)\right] \\ &\geq \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{C}(p \mid \mathscr{E})\right]\end{aligned}$$

and therefore that

$$\operatorname{Poss}(q \mid p) \circledast \operatorname{Poss}(p) \geq \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{C}(p \mid \mathscr{E})\right] \circledast \mathrm{C}(p \mid \mathscr{E})$$

Taking now, in the above expression, the supremum with respect to all propositions p in $\mathscr{P}$, it is

$$\sup_{\mathscr{P}} \left[\operatorname{Poss}(q \mid p) \circledast \operatorname{Poss}(p)\right] \geq$$
$$\sup_{\mathscr{P}} \left\{ \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{C}(p \mid \mathscr{E})\right] \circledast \mathrm{C}(p \mid \mathscr{E}) \right\} \quad (1)$$

Note, however, that since $\mathscr{P}$ is a partition, there always exists a proposition $\hat{p}$ in $\mathscr{P}$ such that $\mathrm{C}(\hat{p} \mid \mathscr{E}) = 1$ (i.e., $\hat{p}$ "intersects" $\mathscr{E}$), and therefore

$$\begin{aligned}&\sup_{\mathscr{P}} \left\{ \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{C}(p \mid \mathscr{E})\right] \circledast \mathrm{C}(p \mid \mathscr{E}) \right\} \\ &\quad \geq \sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(q \mid w) \oslash \mathrm{C}(\hat{p} \mid \mathscr{E})\right] \circledast \mathrm{C}(\hat{p} \mid \mathscr{E}) \\ &\quad = \sup_{w \vdash \mathscr{E}} \mathrm{I}(q \mid w) \\ &\quad = \mathrm{C}(q \mid \mathscr{E})\end{aligned} \quad (2)$$

Theorem 4 follows at once by combination of the inequalities (1) and (2). ■

Finally, notice also that, although Theorems 3 and 4 have been characterized as duals, it is not necessary that $\mathscr{P}$ be a partition for the generalized modus ponens for necessities to hold, although the proof of its possibilistic counterpart relies on such an assumption. It should be clear, however, that richer proposi-

tional collections $\mathscr{P}$ would lead to better lower bounds for values of the degree of implication $\mathbf{I}(q \mid \mathscr{E})$.

Variables

The $\circledast$-transitivity property of $\mathbf{I}$ is the essential fact expressing the relationships between the degrees of implication of the propositions that were proved in the previous section. The statements of these relations in most works devoted to fuzzy logic are made, however, using special subsets of the universe of discourse that are described through the important notion of *variable*. Introduction of this concept, which is also central to other approximate reasoning methodologies, permits us to make a clearer distinction between similarities defined, in some absolute sense, from the several viewpoints and related proximity measures that compare objects (in our case, possible worlds) from the marginal viewpoint of one or more variables.

In what follows, we will assume that only certain propositions, specifying the value of a system variable belonging to a finite set

$$\mathscr{V} = \{X, Y, Z, \ldots\}$$

will be used to characterize possible worlds.

The propositions of interest are those formed by logical combination of statements of the type

"The value of the variable V is v."

where V is in the variable set $\mathscr{V}$ and v is a specific value in the domain $\mathscr{D}(V)$ of the variable V.

We will also assume that, in any possible world, the value of any variable is a member of the corresponding domain of definition of the variable. In the context of our discussion, we will not need to make special assumptions about the scalar or numeric nature of the state variables, using the notion in the same primitive and general sense in which it is customarily used in predicate calculus.

We will be specially interested in subsets, called *variable sets*, of the universe $\mathscr{U}$ consisting of worlds where the value of some variable V is equal to a specified value v. We will denote by $[X = x]$ (similarly $[Y = y]$, etc.) the set of all possible worlds where the proposition "The value of the variable X is x" is true. Clearly, the variable-sets in the collection

$$\{[X = x] : x \text{ is in } \mathscr{D}(X)\}$$

partition the universe into disjoint subsets. These collections have been used to characterize the concept of *rough sets* (Pavlak [30]), of importance in many information system analysis problems, including some that arise in the context

of approximate reasoning. A similar notion has been used also to describe algorithms for the combination of probabilities and of belief functions (Shafer et al. [39]).

To simplify the notation we will write

$$w \vdash x, \quad w \vdash y, \quad \ldots$$

as shorthand for $w \vdash [X = x]$, $w \vdash [Y = y], \ldots,$ respectively.

POSSIBILISTIC STRUCTURES AND LAWS The usual statements of the laws of fuzzy logic are made, as mentioned before, through the use of variables rather than by means of general propositional expressions. It is customary, for example, to speak of the possibility of the variable X taking the value x to describe the value that a possibility function for an evidential set $\mathscr{E}$ attains for the proposition $[X = x]$.

In our model, we will therefore say that a function

$$\text{Poss}(\cdot) : \mathscr{D}(X) \to [0, 1]$$

is a possibility function for the evidential set $\mathscr{E}$ and the variable X whenever

$$\text{Poss}(x) \geq \mathbf{C}([X = x] \mid \mathscr{E})$$

for all values x in the domain $\mathscr{D}(X)$. Similarly, we will say that Nec$(\cdot)$ is a necessity function for X whenever

$$\text{Nec}(x) \leq \mathbf{I}([X = x] \mid \mathscr{E})$$

for all values x in $\mathscr{D}(X)$.

If possibility distributions are defined in this way as point functions in the variable domain $\mathscr{D}(X)$, then it is possible to use the disjunctive laws of fuzzy logic proved in the section Properties of Possibility and Necessity Functions to extend their definition over the power set of $\mathscr{D}(X)$, that is,

$$\text{Nec}(A \cup B) = \max[\text{Nec}(A), \text{Nec}(B)]$$

$$\text{Poss}(A \cup B) = \max[\text{Poss}(A), \text{Poss}(B)]$$

where A and B are subsets of the domain $\mathscr{D}(X)$. These equations are usually given as the basic disjunctive laws of possibility distributions.

Note that, using such extensions, both possibility and necessity functions are nondecreasing functions (with respect to the order induced by set inclusion). The value of Nec(A) measures the extent to which the evidence supports the statement that the variable value necessarily lies in the subset A of its domain of definition, with a dual interpretation being applicable for possibility distributions.

MARGINAL AND JOINT POSSIBILITIES The original similarity relation introduced earlier may be considered to be a measure of proximity between possible worlds from the joint viewpoint of all system variables. The notion of variable, however, permits the definition of similarities from the restricted viewpoint of some variables or subsets of variables.

These restricted perspectives play a role with respect to the original similarity S that is analogous to that of marginal probability distributions with respect to joint probability distributions. To derive useful expressions that describe similarities between two values x and x' of the same variable X, it should be noted first that the degree of implication $\mathrm{I}(\cdot \mid \cdot)$ is transitive. This fact permits the application of a theorem of Valverde [45] to define a function S_X by means of the expression

$$S_X : \mathscr{D}(X) \times \mathscr{D}(X) \to [0,1] : (x, x') \to \min[\mathrm{I}(x \mid x'), \mathrm{I}(x' \mid x)]$$

Defined in this way as a "symmetrization" of the *preorder* induced by the degree of implication $\mathrm{I}(\cdot \mid \cdot)$, the marginal similarity S_X has the properties of a similarity function. Furthermore, the "projection" operation entailed by the use of $\mathrm{I}(x \mid x')$, based on the projection of every x'-world into the set of x-worlds, may be considered to be the basic mechanism to transform the original similarity function into one that discerns differences only in the values of the variable X.

It must be noted, however, that unless additional assumptions are made about the nature of the original similarity S, the function S_X fails to satisfy the intuitive requirement

$$S(w, w') \le S_X(w, w')$$

whenever $w \vdash x$ and $w' \vdash x'$, that is, the similarity between two objects from a restricted viewpoint is always higher than their similarity from more general viewpoints that encompass additional criteria of comparison.

Although considerable research remains to be done to identify alternative definitions of marginal similarities that are not hampered by this problem, a basic result of Valverde [45] presented later in this paper, appears to provide the essential tool that must be employed to produce the required coarser measures. Additional reasonable assumptions that might be demanded from S to facilitate the construction of marginal similarities with desirable characteristics are also an object of current investigation.

CONDITIONAL DISTRIBUTIONS AND GENERALIZED INFERENCE The basic conditional structures of fuzzy logic are usually defined as elastic constraints that restrict the values of one variable given those of another. By simple extension of our previous convention to conditional structures, we will write $\mathrm{Nec}(y \mid x)$ and $\mathrm{Poss}(y \mid x)$ as shorthand for

$$\mathrm{Nec}([Y = y] \mid [X = x]) \qquad \text{and} \qquad \mathrm{Poss}([Y = y] \mid [X = x])$$

respectively.

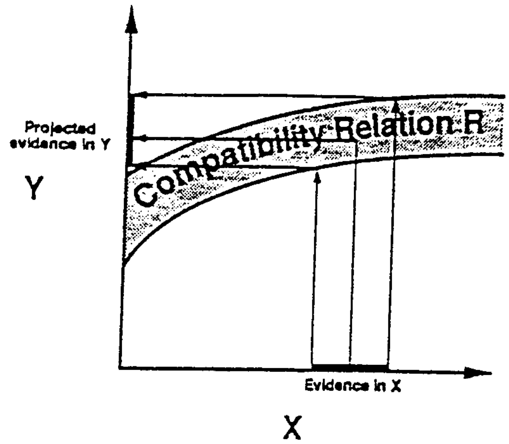


Figure 5. Inference as a compatibility relation.

If a classical (i.e., Boolean) inferential rule of the type

$$\text{If } X = x, \text{ then } Y \text{ is in } R(x).$$

is thought of as the definition of a relation R defined over pairs (x, y) in the Cartesian product $X \times Y$, then such a relation may be used to define a multivalued mapping that maps possible values of X into possible values of Y as illustrated in Figure 5.

Such a *compatibility relation* perspective is an essential element of the original formulations of both the Dempster–Shafer calculus of evidence (Dempster [8]), where distributions in some space (i.e., the domain of some variable X) are mapped into distributions of another variable (i.e., the domain of another variable Y) by direct transfer of "mass" from individual values to their mapped projections, and of the compositional rule of inference (Zadeh [51]).

Note that whenever $\text{Poss}(y \mid x) = 1$, if the bound is actually attained, that is, if

$$\sup_{w \vdash \mathscr{E}} \left[\mathrm{I}(y \mid w) \oslash \mathrm{I}(x \mid w) \right] = 1$$

then it is possible for an evidential world w in $[X = x]$ [i.e., $\mathrm{I}(x \mid w) = 1$] to be such that $w \vdash y$. Pairs (x, y) such that $\text{Poss}(y \mid x) = 1$ may be considered to approximate the *core*[9] of a generalized inferential relation that allows us to determine bounds for the similarity between evidential worlds and those in the variable set $[Y = y]$ on the basis of knowledge of similar bounds applicable to the variable set $[X = x]$. This relation, which is the fuzzy extension of the classical compatibility mapping R illustrated in Figure 5, may be thought of as a descriptor of the behavior, for x-worlds, of the values of the variable Y

[9]The core of a fuzzy set μ: $\mathscr{U} \to [0, 1]$ is the set of all points w such that $\mu(w) = 1$, that is, the points that "fully" belong to μ.

"near" R. The compatibility relation is itself approximated by (or embedded in) the core of the conditional possibility distribution, that is, worlds w such that $w \vdash x$ and $w \vdash y$, and such that $\mathrm{Poss}(y \mid x) = 1$.

Since the collection of the sets $[X = x]$ partitions the universe $\mathcal{U}$ into disjoint sets, then the generalized modus ponens laws can be readily stated in terms of variable values as

$$\mathrm{Nec}(y) = \sup_x \left[\mathrm{Nec}(y \mid x) \circledast \mathrm{Nec}(x)\right]$$

$$\mathrm{Poss}(y) = \sup_x \left[\mathrm{Poss}(y \mid x) \circledast \mathrm{Poss}(x)\right]$$

which clearly shows the basic nature of inferential mapping as the composition of relational combination (i.e., $\circledast$-"intersection") and projection (i.e., maximization).

FUZZY IMPLICATION RULES We will now examine proposed interpretations for conditional rules, usually stated in the form

If X is A, then Y is B.

within the context of possibilistic logic. Whereas in two-valued logic any such rule simply states that whenever a condition A is true, another condition B also holds, various interpretations have been proposed for rules expressing other notions of conditional truth.

In the case of probabilities, for example, degrees of conditionality have been modeled either by means of conditional probability values $\mathrm{Prob}(A \mid B)$, which measure the likelihood of B given the assumed truth of A, or by the alternative interpretation $\mathrm{Prob}(\neg A \vee B)$, used by Nilsson [29] in his probabilistic logic, which essentially quantifies the probability that a rule is a valid component of a knowledge base. Either one of these interpretations is valid in particular contexts being, respectively, the probabilistic extensions of the so-called *de re*, that is,

$$p \rightarrow \Pi q$$

and *de dicto*, that is,

$$\Pi(p \rightarrow q)$$

interpretations of conditionals in modal logic.

In fuzzy logic, two major interpretations have been advanced to translate conditional rules,[10] with A and B corresponding to the fuzzy sets

$$\mu_A : X \rightarrow [0, 1] \qquad \text{and} \qquad \mu_B : Y \rightarrow [0, 1]$$

[10] A rather encompassing account of potential fuzzy reasoning mechanisms may be found in a paper by Mizumoto et al. [27].

The first interpretation was originally proposed by Zadeh [52], as a formal translation of the statement

If μ_A is a possibility for X, then μ_B is a possibility distribution for Y.

This conditional statement, which may be regarded as a constraint on the values of one variable given those of another, states the existence of a conditional possibility function Poss$(\cdot \mid \cdot)$ such that

$$\mu_B(y) \geq \sup_x \left[\operatorname{Poss}(y \mid x) \circledast \mu_A(x)\right] \geq \operatorname{Poss}(y \mid x) \circledast \mu_A(x)$$

Recalling now the definition and properties of the pseudoinverse, we may restate this particular interpretation as

$$\operatorname{Poss}(y \mid x) = \mu_B(y) \oslash \mu_A(x) \geq \mathrm{I}(y \mid w) \oslash \mathrm{I}(x \mid w)$$

for every world $w \vdash \mathscr{E}$.

In Zadeh's original formulation, made within the context of a calculus based on the minimum function as the T-norm, conditionals were, however, formally translated by means of the pseudoinverse of the Lukasiewicz T-norm. Certain formal problems associated with such a combination were pointed out by Trillas and Valverde [42], who developed translations consistent with the T-norm used as the basis for the possibilistic calculus.

Using the characterization of conditionals introduced earlier, this relation may also be thought of as a measure of the degree to which a possibility for Y exceeds a fraction (measured by the conditional possibility distribution) of a given possibility distribution for X. In particular, whenever Poss$(y \mid x) = 1$, then $\mu_B(y) \geq \mu_A(x)$, indicating the *possible* existence, since Poss$(y \mid x)$ is only an upper bound of $\mathrm{I}(y \mid w) \oslash \mathrm{I}(x \mid w)$, of an evidential world such that $w \vdash x$ and $w \vdash y$, with x in A and y in B.

As illustrated in Figure 6, where it has been assumed that the underlying metric (i.e., dissimilarity) is proportional to the Euclidean distance in the plane, the core of the corresponding conditional possibility distribution is an (upper) approximant of a classical compatibility relation (indicated by the shaded area in the figure) that fans outward from the Cartesian product of the cores of A and B. If this interpretation is taken whenever several such rules are available, then each one of these rules will lead to a separate possibility distribution. Combination of these upper bounds by minimization results in a sharper possibility estimate that represents the "integrated" effect of the rule set.

The second interpretation of conditional relations, leading to a wide variety of practical applications (Sugeno [41]), was utilized by Mamdani and Assilian [26] to develop fuzzy controllers. The basic idea underlying this explanation follows an approach originally outlined by Zadeh [47, 48, 49, 50, 51]. In this

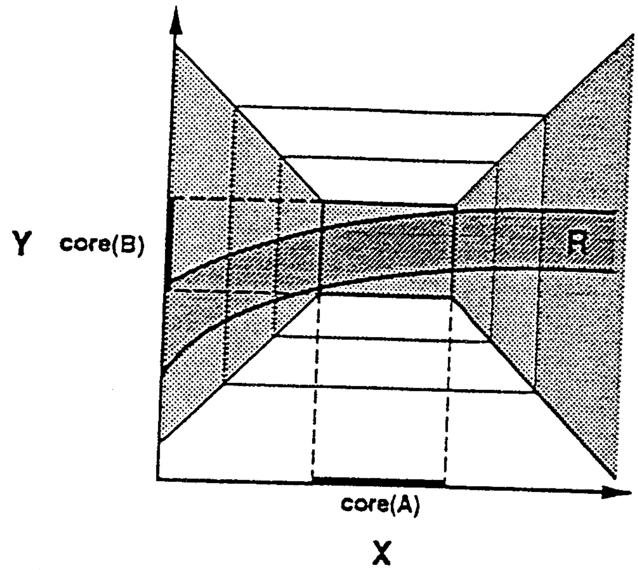


Figure 6. Rules as possibilistic approximants of a compatibility relation.

case, a number of conditional statements of the form

$$\text{If } X \text{ is } A_k, \text{ then } Y \text{ is } B_k, \qquad k = 1, 2, \ldots, n$$

are given as a combined "disjunctive" description of the relation between X and Y, rather than as a set of independently valid rules. The purpose of this rule set is the approximation of the compatibility relation by a "fuzzy curve" generated by disjunction of all the rules in the set, as shown in Figure 7.

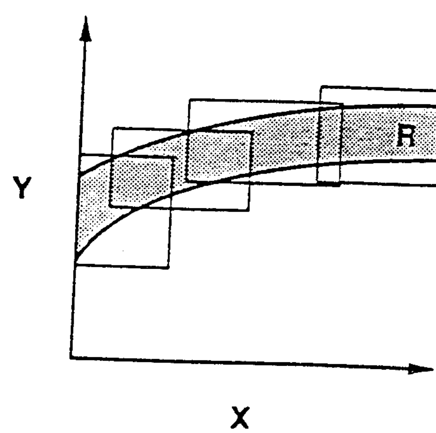


Figure 7. Rule sets as disjunct approximants of a compatibility relation.

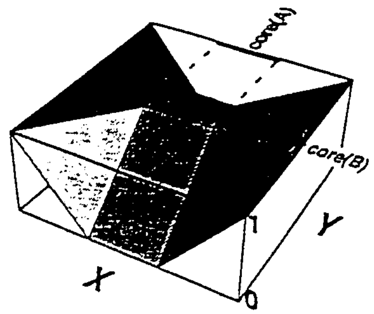


Figure 8. A possibilistic conditional rule (ZTV).

Recalling the characterization of conditioning as an extension of a classical compatibility relation, we may say that the core of the compatibility relation is approximated by above by the union

$$\bigcup_{k=1}^{n}\left[\mathrm{core}(\mu_{A_k}) \times \mathrm{core}(\mu_{B_k})\right]$$

of the Cartesian products of the cores of the fuzzy sets for A_k and B_k. In this case the multiple rules are meant to approximate some region of possible (X, Y) values, and the results of application of individual component rules must be combined using maximization to produce a conditional possibility function. We may say, therefore, that under the Zadeh–Mamdani–Assilian (ZMA) interpretation, the function

$$\mathrm{Poss}(y \mid x) = \sup_k \left\{\min\left[\mu_A(x), \mu_B(y)\right]\right\}$$

is a conditional possibility for Y given X.

It is important to note that the two interpretations of fuzzy rules that we have just examined are based on different approaches to the approximation (by above) of the value

$$\sup_{w \vdash \mathscr{E}}\left[\mathrm{I}(y \mid w) \oslash \mathrm{I}(x \mid w)\right]$$

being, in the case of the Zadeh-Trillas-Valverde (ZTV) method, the result of the *conjunction* of multiple fuzzy relations such as that illustrated in Figure 8, while in the case of the ZMA logic the construction requires *disjunction* of relations such as that illustrated in Figure 9.

The difference between the two approaches when combining several rules is illustrated also in Figures 10 and 11, showing the contour plots for the α-cuts of the fuzzy relations that are obtained in a simple example involving four rules. In these figures, the rectangles with a dark outline correspond to the Cartesian products of the cores of the antecedents A_k and B_k. Darker shades of gray correspond to higher degrees of membership.

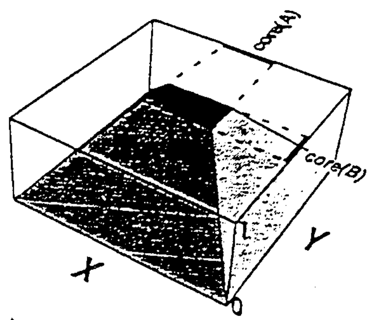


Figure 9. A component of a disjunctive rule set (ZMA).

The reader should be cautioned, however, about the potential for invalid comparisons that may result from hasty examination of these figures. Each formalism should be regarded as a procedure for the approximation of a compatibility relation that is based on a different approach for the description of relationships between variables. In the case of the ZMA interpretation, the intent is to generalize the interpolation procedures that are normally employed in functional approximation. As such, this approach may be said to be inspired by the methodology of classical system analysis. The ZTV approach, by contrast, is a generalization of classical logical formulations and may be regarded, from a relational viewpoint, as a procedure to describe a function as the locus of points that satisfies a set of constraints rather than as a subset of "fuzzy points" of a Cartesian product.

Figures 10 and 11, while showing that the same rule sets would lead to radically different results, should not be considered, therefore, to discredit interpolative approaches, as such techniques, proceeding from a different

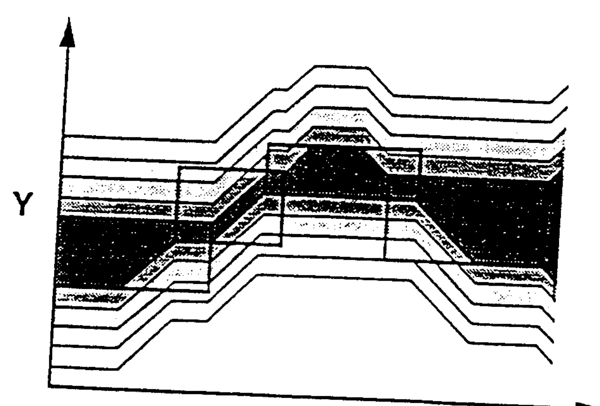


Figure 10. Contour plots for a rule set (ZTV).

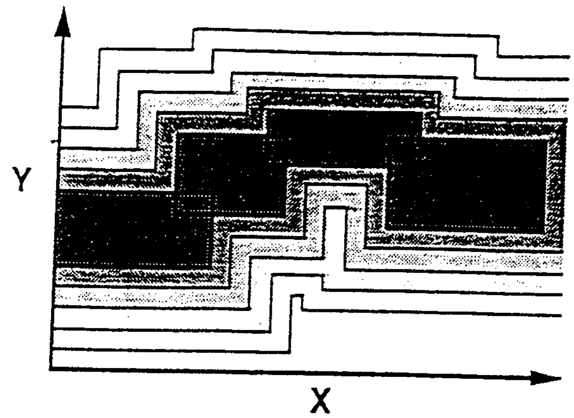


Figure 11. Contour plots for a rule set (ZMA).

perspective, should normally be based on rule sets that are different from those used when rules are thought of as independent constraints.

THE NATURE OF SIMILARITY RELATIONS

In this closing section, we will examine issues that arise naturally from our previous examination of the role of similarities as the semantic basis for possibility theory.

Our discussion focuses on two topics. We look first at the requirements that our theory imposes upon the nature of the scales used to measure proximity or resemblance between possible worlds. Finally, our examination of the interplay between similarities and possibilities turns to issues related to the generation of similarity relations from such sources as domain knowledge that describe significant relations between system variables.

On Similarity Scales

Our previous interpretation of possibilistic concepts and structures was based on the use of measures of proximity that quantify interobject resemblance using real numbers between 0 and 1. Our assumptions about the use of the [0, 1] interval as a similarity scale have been made primarily, however, as a matter of convenience to simplify the description of our model while being consistent with the customary definitions of possibility and necessity distributions as functions taking values in that interval.

Close examination of the actual requirements imposed upon our similarity scales reveals, however, that our measurement domain may be quite general so as to include symbolic structures such as

$$\{\mathit{identical}, \mathit{very\ similar}, \ldots, \mathit{completely\ dissimilar}\}$$

Our model is based on the use of a partially ordered set having a maximal anc a minimal element representing identity and complete dissimilarity, respectively. Furthermore, we have assumed the existence of a binary operation (the triangular norm ⊛) mapping pairs of possible worlds into real numbers, with certain desirable order-preserving and transitive properties. The concept of triangular norm, however, does not rely substantially on the use of real numbers as its range and may be readily extended to more general partially ordered sets with maximal and minimal elements.

We have also assumed a continuity property for the triangular norm operation. This property, however, simply requires that a notion of proximity also exist among similarity values so as to provide a form of (order-consistent) topology in that space. While, in general, more precise scales will result in more detailed representations of interworld similarity, it is important to stress that the similarity-based model presented here does not rely on "density" assumptions such as the existence of an intermediate value c between any different values a and b in the similarity-measurement scale.

From a practical viewpoint, the major requirement is to quantify proximity in such a way as to be able to determine that two quantities are similar to some degree (i.e., approximate matching). The degree of precision that such a matching entails is problem-dependent and will typically be the result of conflicting impositions between the desire, on the one hand, to keep granularity relatively low to reduce complexity, and the need, on the other, to describe system behavior at an acceptable level of accuracy. The work of Bonissone and Decker [4] is a significant example of the type of systematic study that must be carried out to define similarity scales that are both useful and tractable.

The Origin of Similarity Functions

The model of fuzzy logic presented in this paper is centered on the metric notion of similarity as a primitive concept that is useful in explaining the nature of possibilistic constructs and the meaning of possibilistic reasoning. In this formulation, similarities are defined as real functions defined over pairs of possible worlds.

From this perspective, similarities describe relations of resemblance between objects of high complexity, which, typically, result from consideration of a large number of system variables. Reliance on such complex structures has been the direct consequence of a research program that stressed conceptual clarification as its primary objective. In practice, however, it will be generally difficult to define complex measures that quantify similarity between complex objects on the basis of a large number of criteria.

Similarities provide the framework that is required to understand approximate relations of corelevance, usually stated as generalized conditional rules. The practical generation of similarity functions typically proceeds, however, in

the opposite direction, from separate statements about limited aspects of system behavior to general metric structures. Once such resemblance measures are defined, they may be used to express and acquire new laws of system behavior determined, for example, from historical experience with similar systems. Furthermore, such similarity notions may be used as the basis for analogical reasoning systems that try to determine the system's state on the basis of similarity to known cases (Kolodner [23]).

Perhaps the simplest mechanism that may be devised to generate complex metrics from simpler ones is that which starts with measures of resemblance that quantify proximity from a limited viewpoint. These metrics are usually derived, using a variety of techniques, in unsupervised pattern classification (or clustering) problems (Hartigan [20]). In many important applications, hierarchical taxonomies—a feature of many representation approaches in artificial intelligence—may be used, often in connection with a variety of weighing schemes, quantifying branching importance, to generate metrics that often satisfy the more stringent requirements of an ultrametric (Jardine and Sibson [22]).

Classification hierarchies such as those may be though of as sets of general rules, having a particularly useful structure, that specify interest proximity from relevant, but restricted, viewpoints, eventually providing measures of similarity between variable values (i.e., the "leaves" of the taxonomic tree). More generally, however, we may expect that sets of possibilistic rules (i.e., a general knowledge base) defining a general semantic network of corelevance relations may be available as the source for the determination of interobject proximity. These possibilistic semantic networks resemble conventional semantic networks in most regards, being more general in that, in addition to specifying knowledge about system behavior in some subsets of state-space,[11] they also specify characteristics of behavior in neighborhoods of those subsets.

We may think, therefore, that the antecedents of implicational rules define general regions in state-space where existence of relevant knowledge may increase insight through application of inferential rules. Using Zadeh's terminology, these antecedents define "granules" that identify important regions of state-space and indicate the level of accuracy (or *granularity*) that is required to perform effective system analysis. In this case, the possibilistic granules correspond to fuzzy sets that are used to specify both what is true in the core of the granule and, with decreasing specificity, what is true in a nested set (i.e., the α-cuts) of its neighborhoods. The ability to specify behavior using such a topological structure results in inferential gains that are the direct consequence of our ability to reason by similarity—an ability that is made possible by the approximate matching property of the generalized modus ponens. From an-

[11] The expression "state-space" is loosely used here to indicate the space defined by all system variables.

other perspective yet, the fuzzy granules identified by possibilistic rules may also be thought of as generalizations of the arbitrary variable sets used in a variety of artificial intelligence efforts aimed at understanding system behavior using qualitative descriptions of reality (Forbus [16]).

A number of heuristics may be easily formulated to integrate "marginal" measures of resemblance into joint similarity relations. More generally, however, we may state the problem of similarity construction as that of defining metric structures on the basis of knowledge of the aspects of system behavior that are important to its understanding—the previously mentioned granules, which define what must be distinguished. Since generally those granules are fuzzy sets, the relevance to similarity construction of the following representation theorem, due to Valverde, may be immediately seen.

THEOREM 5 (VALVERDE)) *A binary function S mapping pairs of objects of a universe of discourse $\mathscr{U}$ into* $[0,1]$ *is a similarity relation if and only if there exists a family $\mathscr{H}$ of fuzzy subsets of $\mathscr{U}$ such that*

$$S(w, w') = \inf_{\mathscr{H}} \{\min[h(w) \oslash h(w'), h(w') \oslash h(w)]\}$$

for all w and w' in $\mathscr{U}$, where the infimum is taken over all fuzzy subsets h in the family $\mathscr{H}$.

Besides its obvious relevance to the generation of similarity relations from knowledge of important sets in the domain of discourse, Valverde's theorem—resulting originally from studies in pattern recognition—is also of potential significance to the solution of knowledge acquisition problems because of the important relations that exist between learning procedures and structure-discovery techniques such as cluster analysis.

CONCLUSION

This paper has presented a similarity-based model that provides a clear interpretation of the major structures and methods of possibilistic logic using metric concepts that are formally different from the set-measure constructs of probability theory. Regardless of the potential existence, so far unestablished, of probability-based interpretations for possibilistic structures, this metric model makes clear that there are no compelling reasons to confuse two rather different aspects of uncertainty into a single notion simply because one's favorite theoretical framework, in spite of its otherwise many remarkable virtues, fails to fully capture reality.

Succinctly stated, being in a situation that resembles a state of affairs S does not make S likely or vice versa. Furthermore, our reference state may not even be possible in the current circumstances, which would make it completely unlikely, but we may still find it useful as a comparison landmark. This use of

"impossible" examples as a way to illustrate system behavior is very prevalent in human culture, being exemplified by such utterances as "he had the strength of a horse and the swiftness of a swallow," even if it is obvious to all that no such beast exists other than for such metaphorical purposes.

The insight provided by this model makes it rather obvious that very little can be gained by continuing to assert a potential—although never revealed—encompassing probabilistic interpretation for possibilistic structures that, presumably, would render them unnecessary as serious objects of scientific discourse. In addition, and quite beyond whatever understanding theory may provide, the current success of possibilistic logic as the basis for major systems of important human value (Sugeno [41]), often unmatched by other approaches, should be enough to convince those having more pragmatic perspectives as to its utility.

The task for approximate reasoning researchers is to proceed now beyond unnecessary controversy into the study of the issues that arise from models such as the one presented in this paper. Among such questions, further studies of the relations between the notions of possibility, similarity, and negation and of those between probability and possibility are of major importance.

ACKNOWLEDGMENTS

This work was supported in part by the Air Force Office of Scientific Research under contract No. F49620-89-0001 and in part by the U.S. Army Research Office under contract No. DAAL03-89-K-0156. The views, opinions, and/or conclusions contained in this paper are my own and should not be interpreted as representative of the official positions, decisions, or policies, either express or implied, of the Air Force Office of Scientific Research, the Department of the Army, or the U.S. Government.

The model presented herein is the product of an effort aimed at the explication of possibilistic structures by means of more primitive similarity concepts. During this endeavor, which produced sufficient versions to require eventually a formal numbering system, I benefited from the advice and comments of Claudi Alsina, Hamid Berenji, Piero Bonissone, Didier Dubois, Francesc Esteva, Oscar Firschein, Marty Fischler, Pascal Fua, María Angeles Gil, Luis Godo, Andy Hanson, Jerry Hobbs, David Israel, Joan Jacas, Yvan Leclerc, Ramón López de Mántaras, John Lowrance, Abe Mamdani, Bob Moore, Ray Perrault, Henri Prade, Elie Sanchez, Philippe Smets, Tom Strat, Enric Trillas, Llorenç Valverde, Len Wesley, and Lotfi Zadeh. To all of them many thanks.

The Fulbright Commission for the International Exchange of Scholars, the University of the Balearic Islands, the Center of Advanced Studies of Blanes, the National Research Council of Spain, and the Caixa de Pensions of

Barcelona supported several visits to Spanish research centers where I had invaluable exchanges on the subject matter of this paper.

The assistance of Joani Ichiki, Valerie Maslak, and Diego Ruspini in the preparation of the final manuscript is gratefully acknowledged.

References

1. Alsina, C., and Trillas, E., Additive homogeneity of logical connectives for membership functions, in *Analysis of Fuzzy Information*, Vol. I, *Mathematics and Logic* (J. C. Bezdek, Ed.), CRC, Boca Raton, Fla., 1987, pp. 179–184.

2. Baldwin, J. F., A new approach to approximate reasoning using a fuzzy logic, *Fuzzy Sets Syst.* 2, 302–335, 1979.

3. Bezdek, J. C., and Harris, J. O., Fuzzy partitions and relations: an axiomatic basis for clustering, *Fuzzy Sets Syst.* 1, 112–127, 1978.

4. Bonissone, P., and Decker, K., Selecting uncertainty calculi and granularity: an experiment in trading-off precision and uncertainty, in *Uncertainty in Artificial Intelligence* (L. N. Kanaal and J. F. Lemmer, Eds.), North-Holland, Amsterdam, 1986.

5. Bonissone, P. P., Gans, S. S., and Decker, K. S., RUM: a layered architecture for reasoning with uncertainty, in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, (John McDermott, Ed.), Morgan Kaufmann, Los Altos, Calif., 1987, pp. 891–896.

6. Carnap, R., *The Logical Foundations of Probability*, Univ. Chicago Press, Chicago, 1950.

7. Choquet, G., Théorie des capacités, *Ann. Inst. Fourier (Grenoble)* V, 131–295, 1953.

8. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Stat.* 38, 325–339, 1967.

9. Dieudonné, J., *Foundations of Modern Analysis*, Academic, New York, 1960.

10. Dubois, D., and Prade, H., Fuzzy logics and the generalized modus ponens revisited, *Int. J. Cybern. Syst.*, 15, 293–331, 1984.

11. Dubois, D., and Prade, H., *Possibility Theory: An Approach to the Computerized Processing of Uncertainty*, Plenum, New York, 1988.

12. Dubois, D., and Prade, H., In search of a modal system for possibility theory, *Proceedings of the 8th European Conference on AI*, Technical University, Munich, 1988, pp. 501–506.

13. Dubois, D., and Prade, H., An introduction to possibility theory and fuzzy logics, in *Non-standard Logics for Automated Reasoning* (P. Smets, E. H. Mamdani, D. Dubois, and H. Prade, Eds.), Academic, New York, 1988, pp. 287–326.

14. Dubois, D., and Prade, H., Representation and combination of uncertainty with belief and possibility measures, *Comput. Intell.*, to appear.

15. Dubois, D., and Prade, H., Rough fuzzy sets and fuzzy rough sets, *Int. J. Gen. Syst.*, to appear.

16. Forbus, K., Qualitative process theory, *AI*, 24, 85–168, 1984.

17. Gaines, B., Fuzzy and probability uncertainty logics, *Inf. Control* 38, 154–169, 1978.

18. Godo, L. I., López de Mántaras, R., Sierra, C., and Verdaguer, A., Managing linguistically expressed uncertainty in MILORD: application to medical diagnosis, Res. Report No. 87/2, Group of Logic and Artificial Intelligence, Center of Advanced Studies of Blanes, Blanes, Spain, 1987.

19. Halmos, P. R., *Measure Theory*, Springer-Verlag, New York, 1974.

20. Hartigan, J., *Clustering Algorithms*, Wiley, New York, 1975.

21. Hughes, G. E., and Creswell, M. J., *An Introduction to Modal Logic*, Methuen, New York, 1972.

22. Jardine, N., and Sibson, R., *Mathematical Taxonomy*, Wiley, New York, 1971.

23. Kolodner, J. (Ed.), *Case-Based Reasoning*, Proceedings of a Workshop, Clearwater Beach, Fla., May 1988, Morgan Kaufmann, San Mateo, Calif., 1988.

24. Lakoff, G., Hedges: a study in meaning criteria and the logic of fuzzy concepts, *J. Phil. Logic* 2, 458–508, 1973.

25. Lewis, D., *Counterfactuals*, Harvard Univ. Press, Cambridge, Mass., 1973.

26. Mamdani, E. H., and Assilian, S., An experiment in linguistic synthesis with a fuzzy logic controller, *Int. J. Man-Mach. Stud.* 7, 1–13, 1975.

27. Mizumoto, M., Fukami, S., and Tanaka, K., Some methods of fuzzy reasoning, in *Advances in Fuzzy Set Theory and Applications* (M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds.), North-Holland, Amsterdam, 1979, pp. 117–136.

28. Murai, T., Miyakoshi, M., and Shimbo, M., Fuzzifications of modal operators from the standpoint of fuzzy semantics, *Proceedings of the 2nd International Fuzzy Systems Assoc. Congress*, Tokyo, 430–432, 1987.

29. Nilsson, N. J., Probabilistic logic, *AI* 28, 71–87, 1987.

30. Pavlak, Z., Rough sets, *Int. J. Comput. Inf. Sci.* 11, 341–356, 1982.

31. Rescher, N., *Many Valued Logic*, McGraw-Hill, New York, 1969.

32. Ruspini, E. H., A new approach to clustering, *Inf. Control* 15, 22–32, 1969.

33. Ruspini, E. H., A theory of cluster analysis, Ph.D. Thesis, Dept. of System Science, Univ. California, Los Angeles, 1977.

34. Ruspini, E. H., Recent developments in fuzzy clustering, in *Fuzzy Sets and Possibility Theory: Recent Developments* (R. R. Yager, Ed.), Pergamon, New York, 1982, pp. 133–147.

35. Ruspini, E. H., The logical foundations of evidential reasoning, Tech. Note No. 408, AI Center, SRI International, Menlo Park, Calif., 1987.

36. Schocht, P. K., Fuzzy modal logic, *Proceedings of the 5th International Symposium on Multiple-Valued Logic*, IEEE, 176–182, 1975.

37. Schweizer, B., and Sklar, A., Associative functions and abstract semigroups, *Publ. Math. Debrecen* 10, 69–81, 1963.

38. Shafer, G., *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.

39. Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating belief functions in qualitative Markov trees, *Int. J. Approximate Reasoning* 1, 349–400, 1987.

40. Sokal, R. R., and Sneath, P. H. A., *Principles of Numerical Taxonomy*, Freeman, San Francisco, 1963.

41. Sugeno, M., *Industrial Applications of Fuzzy Control*, North-Holland, Amsterdam, 1985.

42. Trillas, E., and Valverde, L., On some functionally expressible implications for fuzzy set theory, in *Proceedings 3rd International Seminar on Fuzzy Set Theory* (E. P. Klement, Ed.), Johanes Kepler Univ., Linz, Austria, 173–190, 1981.

43. Trillas, E., and Valverde, L., On mode and implication in approximate reasoning, in *Approximate Reasoning and Expert Systems* (M. M. Gupta, A. Kandel, W. Bandler, and J. B. Kiszka, Eds.), North-Holland, Amsterdam, 1985, pp. 157–166.

44. Tversky, A., Features of similarity, *Psycholog. Rev.* 84, 433–460, 1977.

45. Valverde, L., On the structure of F-indistinguishability operators, *Fuzzy Sets Syst.* 17, 313–328, 1985.

46. Zadeh, L. A., Fuzzy sets, *Inf. Control* 8, 338–353, 1965.

47. Zadeh, L. A., A rationale for fuzzy control, *J. Dynam. Syst., Measurement Control* C94, 3–4, 1972.

48. Zadeh, L. A., Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. Syst. Man Cybern.*, SMC-3, 28–44, 1973.

49. Zadeh, L. A., The concept of a linguistic variable and its application to approximate reasoning, Part 1, *Inf. Sci.* 8, 199–249, 1975.

50. Zadeh, L. A., The concept of a linguistic variable and its application to approximate reasoning, Part 2, *Inf. Sci.* 8, 301–357, 1975.

51. Zadeh, L. A., The concept of a linguistic variable and its application to approximate reasoning, Part 3, *Inf. Sci.* 9, 43–80, 1976.

52. Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.* 1, 3–28, 1978.

53. Zadeh, L. A., A theory of approximate reasoning, in *Machine Intelligence*, Vol. 9 (D. Michie and L. I. Mikulich, Eds.), Halstead, New York, 1979, pp. 149–194.

54. Zadeh, L. A., The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets Syst.* 11, 199–227, 1983.

VOLUME 5, NUMBER 1, JANUARY 1991

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

Editor-in-Chief
James C. Bezdek

NORTH-HOLLAND
ISSN: 0888-613X
IJARE4 5(1) 1-88 (1991)

Approximate Reasoning:
Past, Present, Future

APPROXIMATE REASONING: PAST, PRESENT, FUTURE

Technical Note No. 492

June 27, 1990

By: Enrique H. Ruspini, Sr. Computer Scientist
Artificial Intelligence Center
Computing and Engineering Sciences Division

Published in Information Sciences Volume 57-58, pp 297-317 (1991)

This work was supported in part by the Air Force Office of Scientific Research under Contract F49620-89-K-0001 and in part by the United States Army Research Office under Contract No. DAAL03-89-K-0156.

Abstract

This note presents a personal view of the state of the art in the representation and manipulation of imprecise and uncertain information by automated processing systems. To contrast their objectives and characteristics with the sound deductive procedures of classical logic, methodologies developed for that purpose are usually described as relying on Approximate Reasoning.

Using a unified descriptive framework, we will argue that, far from being mere approximations of logically correct procedures, approximate reasoning methods are also sound techniques that describe the properties of a set of conceivable states of a real-world system. This framework, which is based on the logical notion of possible worlds, permits the description of the various approximate reasoning methods and techniques and simplifies their comparison. More importantly, our descriptive model facilitates the understanding of the fundamental conceptual characteristics of the major methodologies.

We examine first the development of approximate reasoning methods from early advances to the present state of the art, commenting also on the technical motivation for the introduction of certain controversial approaches.

Our unifying semantic model is then introduced to explain the formal concepts and structures of the major approximate reasoning methodologies: classical probability calculus, the Dempster-Shafer calculus of evidence, and fuzzy (possibilistic) logic. In particular, we discuss the basic conceptual differences between probabilistic and possibilistic approaches.

Finally, we take a critical look at the controversy about the need and utility for diverse methodologies, and assess requirements for future research and development.

Contents

1 Introduction

This note presents a personal view of the state of the art in approximate reasoning, the name used to describe several methodologies for the development of intelligent systems capable of manipulating imprecise and uncertain information.

Approximate reasoning techniques loosely based on the calculus of probability appeared almost simultaneously with the development of expert systems relying on classical (i.e., two-valued) logic techniques. Soon after these systems were introduced, other approaches to the treatment of uncertainty and imprecision were also proposed, both to generalize more or less conventional probabilistic schemes and to capture other aspects of imperfect knowledge, claimed to have a nonprobabilistic nature.

The short technological history of approximate reasoning methods may be described as being, from that moment, one of extreme controversy that has lasted to this day. Most of the proponents of classical probabilistic treatments, often described, although vaguely and somewhat misleadingly, as Bayesians,[1] have doubted the necessity for the introduction of other conceptual structures and have often sought to explain those frameworks in terms of probabilistic notions. Proponents of alternative approaches, on the other hand, have defended their techniques on the strength of two main arguments: the practical problems associated with the parameter-intensive procedures of conventional probability, often demanding knowledge of a large number of probability values; and, the nonprobabilistic nature of the uncertainties associated with the use of vague concepts.

Much of this disagreement has been clearly caused by misunderstandings about the fundamental philosophical characteristics of each approach. Lacking a suitable basis to interpret certain concepts, particularly those related to the "degrees of truth" of multivalued logics, it has been impossible, until recently, to provide an adequate framework to discuss fundamental issues in a rational manner.

This position paper on the past evolution of the field, its present state of the art, and desiderata for future evolution is the result of recent research by the author in basic semantic issues that are germane to the foundations of approximate reasoning. The presentation is based on the use of a central unifying framework: a formal model of the approximate reasoning problem that explains the similarities and differences between major methodologies. Using this "possible-worlds" model, we will also be able to compare the rationale of nonmonotonic logic approaches with that of approx-

[1]The qualifier Bayesian is used in the context of statistics to describe proponents of a statistical methodology and in the context of the philosophy of probability to denote various subjective views of probability. In Artificial Intelligence, the term has been loosely applied both to those investigating approaches based on the probability calculus and, more narrowly, to those espousing the decision-theoretic methods of subjective probability.

imate reasoning procedures. Although our model is a rigorous formalism, described in detail elsewhere [32,33] in connection with the logical foundations of the Dempster-Shafer calculus of evidence and fuzzy logic, our discussion will be kept as informal as possible to facilitate understanding our philosophical and technical position.

We will contend that regarding probabilistic and possibilistic approaches as competing alternatives is incorrect and confuses the need to describe different aspects of reality with the adequacy or ability of probability as a measure of likelihood. We will also take a critical look at the major claims supporting a narrow view of probability, based on a subjectivist interpretation that regards all forms of rational decision-making as necessarily demanding optimization of expected-utility functionals, and we dispute claims that only such approaches are endowed with either a suitable or a proven decision-theoretical apparatus.

On the basis of our theoretical arguments, and of recent success in the application of various techniques to practical problems, we will also argue that future accomplishment in the field lies in the rational development of tools leading to multiple complementary views of the implications of evidence rather than on arbitrary circumscription to a limited class of techniques and procedures.

2 The Development of Approximate Reasoning

Intelligent systems relying on approximate reasoning techniques [8,39] appeared in the 1970s, approximately at the same time as other systems seeking to emulate the expertise of specialists in diverse fields of endeavor. Problems related to the development of the expert systems based on classical deductive procedures, however, were primarily related to the need to organize knowledge and its processing in such a manner as to assure an efficient derivation of the truth value of hypotheses (i.e., either **true** or **false**). Systems such as MYCIN or PROSPECTOR— reasoning about medical and geological systems, where knowledge is limited and where observations may be difficult or impossible to make—were forced to deal, in addition, with issues that, to this day, have almost completely consumed the attention of approximate reasoning researchers.

These issues may be generally described as related to the extension of the basic derivation rule of classical logic, the modus ponens, which states that from the validity of an antecedent proposition p and that of the implication $p \rightarrow q$, it is possible to derive the validity of the consequent proposition q. Although a conventional expert system, using classical rules of derivation, could be assumed to have sufficient information to derive the validity of a hypothesis of interest, whenever knowledge was scarce or uncertain it was necessary to resort to other schemes that qualified in one way or another the meaning of the truth of propositions. Still imitating the

network-oriented techniques of truth-value propagation of two-valued logic, the approximate reasoning schemes developed in early systems sought to propagate numeric truth values that were loosely related to probabilistic interpretations of uncertainty.

The concept of probability provides a most important tool to describe the state of systems that are known under less than desirable informational circumstances. Arising clearly from the need to make decisions despite undesirable knowledge handicaps, the notion of probability, seriously studied from the seventeenth century, has always played a major role in human judgment [16].

The appeal of probability as an instrument to assess system behavior is due to the empirically observed property that is expressed by the long-run stability of occurrence of certain events. Whether such a pattern of occurrence has been objectively quantified through experimentation or historical observation (objective interpretation), or is subjectively expressed by the willingness to gamble with certain stakes (subjective interpretation), it is clear that it provides a rational basis to formulate rational expectations about system state. Why would anybody, if such predictable stability of occurrence could not be assured, be willing to consciously bet on some outcomes rather than others if the real world defies any attempts to descriptive characterization?

Curiously enough, although probabilistic interpretations were always implicitly or explicitly intended by the developers of early approximate reasoning systems, and while the underlying calculi reflect such explanations, it seems also clear that the machinery of these devices was primarily oriented toward the emulation of the propagation schemes of classical logic with truth flowing from node to node through edges corresponding to implication rules. Approximate truth, measured by numbers associated with objective likelihood or expert confidence, also flowed from evidence to hypothesis in a scheme that generalized the true-false dichotomy of multivalued logic.

Regardless of the clearly intended probabilistic interpretations of those numbers, misgivings about their meaning and utility were sufficient to plant the seeds of the ensuing controversy. Concerns about the inability of probability to capture notions of evidential confirmation led the developers of MYCIN[39], for example, to introduce modified concepts ("certainty factors") as an alternative to direct use of conditional probabilities. In spite of subsequent studies showing that such certainty factors were related to probability values [18], it is clear that these worries were well founded, having been already eloquently expressed in the works of philosophers of science [34].

Although such concerns are indeed important and, despite some claims to the contrary, must still be properly addressed, other issues soon captured the attention of those seeking to develop expert systems with approximate reasoning capabilities. Beyond certain troublesome issues that were apparent when formulating the probabilistic calculi used by PROSPECTOR, arising from inconsistencies between "expert

estimates" of probability values and the laws of probability, it was also clear to those engaged in the development of new expert systems that a typical application required estimation of a very large number of individual probability values [14], which were neither available or derivable from existing data.

In addition, other researchers, acquainted with the concepts and methods of multivalued logic [31,13], advanced the notion that some of the "degrees of truth" being propagated could be interpreted in a nonprobabilistic fashion. The theory of fuzzy sets, introduced by Zadeh in 1965 [45], had been for some time the focus of attention of these researchers and soon became a major source of techniques for the treatment of uncertainty by use of nonprobabilistic schemes.

The variety of approximate reasoning methods arising from this diversity—expressed as a preference toward either a variedly interpreted, more or less strict application of classical probability schemes; as approaches seeking the expression of ignorance about probability values, such as the Dempster-Shafer calculus of evidence; and as nonprobabilistic schemes like fuzzy logic— have led to a controversy that has endured to this day.

It has not been possible, until recently, to discuss these approaches with the help of a unifying framework that facilitates the interpretation of relevant concepts and the comparison of alternative methodologies. This unifying framework is based on a view of approximate reasoning problems as those wherein the truth-value of a hypothesis cannot be deduced from available information.[2] In other words, several scenarios, all consistent with evidence, may be conceived. In some of those stuations the hypothesis is true, while in others it is false.

The logical notion that we will use to characterize such conceivable states of affairs, situations, or scenarios, is the concept of " possible world" utilized by Carnap [4] in his logical treatment of the concept of probability, which was also employed by Nilsson [26] to derive a logic-based methodology for probabilistic reasoning.

3 Possible-World Models

A *possible world* may be briefly described as a function that assigns one and only one of the truth values **true** or **false** to every proposition (i.e., declarative statement) about the system that is being reasoned about. If we seek to describe and study the weather in Menlo Park, for example, the atmospheric conditions at several points in time are described by assigning specific values to meteorological variables such as temperature, humidity, and rainfall, or, equivalently, by assigning a truth value to

[2] Sometimes this characterization is extended to include those cases where that derivation is very difficult.

propositions such as

The temperature at 3PM was 75°F.

Since the value of system variables is unique (e.g., the temperature cannot be both 75°F and 85°F at the same time), it is clear that each possible world (i.e., an assignment of truth values) must satisfy certain consistency conditions that follow from the axioms of classical logic.

In approximate reasoning problems, however, we can usually do more to restrict the extent of the set of possible worlds that may conceivably describe the state of the system. Typically, the information or knowledge about the state of the system and its applicable rules of behavior, in spite of its defficiencies, is a major source of constraints that further limit the extent of the situations that must be considered. The subset of possible worlds that is logically consistent with this evidence is called the *evidential set,* and, in one form or another, is the concern of every approximate reasoning approach. In any approximate reasoning problem, by definition, some of these evidential worlds are such that a hypothesis is true in some of them and false on others, as depicted in Figure 1.

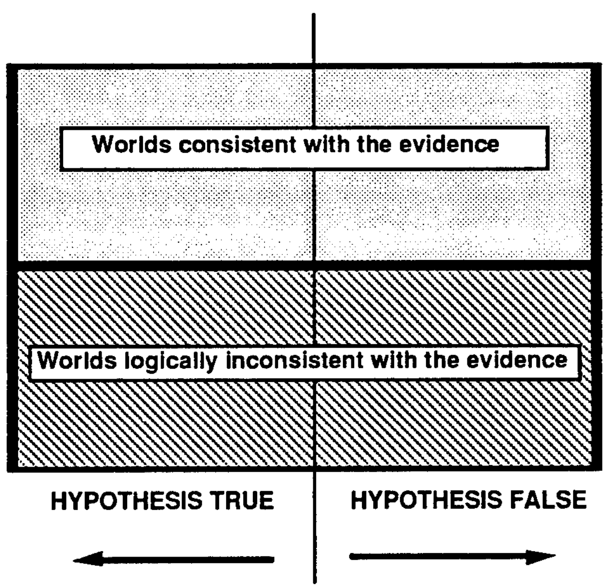


Figure 1: The approximate reasoning problem

The view of approximate reasoning problems that is afforded by this possible-world perspective also simplifies the understanding of the objective of approximate reasoning approaches. Lacking, by the nature of the problem, the ability to determine if the evidence implies whether we are in a situation where a hypothesis is true or in

one where it is false, every approximate reasoning methodology seeks answers to a different problem: that of describing certain properties of the evidential set.

4 The Semantics of Approximate Reasoning

Our view of approximate reasoning methods as techniques to describe the evidential subset[3] e of possible worlds that are consistent with available information now allows a more detailed look into their philosophical bases.

Probabilistic methods, regardless of their subjective or objective semantics, seek to estimate measures of the subsets of the evidential set where a hypothesis h is true and where it is false, i.e., the values

$$\mu(h \wedge e) \quad \text{and} \quad \mu(\neg h \wedge e)\,,$$

or other related quantities, such as likelihood ratios or conditional measures with respect to the evidential set e. The measure μ is, however, an aggregate measure of set extension based on the additive law

$$\mu(p) + \mu(q) = \mu(p \wedge q) + \mu(p \vee q)\,,$$

stating that its value over a set may be derived from knowledge of its value over a partition of nonintersecting subsets. Regardless of the mechanism used to derive the weights associated with individual members of the subsets, it should be clear that interactions and associations between possible worlds (e.g., distances) do not play any role in such quantities. Simply stated, all that matter are the weights of each individual point (more generally, each atomic subset) that are then added to gauge the extent of the subset.

Possibilistic methods, on the other hand, are based on notions of proximity and resemblance between pairs of possible worlds. This association or similarity is also a measure, albeit not one that may be expressed in terms of individual weights. Exploiting the idea that, in many systems, statements that are true in certain situations remain approximately true in similar instances (e.g., clothing that is appropriate when the temperature is 75°F will work nearly as well at 78°F), the purpose of possibilistic techniques is to describe the evidential set in terms of the similarity of its component possible worlds to other possible worlds used as reference landmarks.

The basic difference between probabilistic and possibilistic methods, therefore, goes beyond the use of different formulas to derive truth values. The methodologies are based on different conceptual approaches to the description of the evidential set;

[3]For simplicity, we refer loosely to sets and propositions are if they were the same objects.

they stress, in probabilistic reasoning, relative measures of set size, such as the ratio of previously observed true and false cases, while, in possibilistic reasoning, they stress binary measures of similarity that describe how far is any conceivable scenario from certain significant situations.

In both approaches, however, the objective is the description of properties of the evidential set rather than of any of its particular members. By contrast, certain nonmotonic logic techniques such as circumcription [24] rely on methods to choose least-exceptional worlds in the evidential set by extension of the "close-world assumption" [30], i.e., the only propositions or predicates that are true are those that are known to be true. These techniques may be considered general procedures to represent states of evidential knowledge by choice of prototypical situations. New evidence, however, may force retraction of some of the assumptions leading to the selection of other evidential worlds as prototypes. Another class of nonmonotonic reasoning techniques, while generally fitting the description given above, relies on prespecified "default" rules [29] to control the choice of prototypical worlds. Since these rules are usually formulated on the basis of plausibility notions rooted on statistical information (as in the famous example of Tweety and the flying ability of most live birds) it is not surprising that the derivation techniques and rules of these *preferential logics*—a name indicating their definition of a preferred order for models of a situation—resemble those of probabilistic reasoning. In fact, recent developments strongly point to the existence of a common unifying interpretation for both [28,15].

4.1 Probabilistic Reasoning

There can be little argument from any quarter that frequencies of occurrence of events satisfy the famous additive law that is axiomatized in the definition of set measure [17]. If propositions that describe event occurrence can only be assigned one and only one of the classical probability values, then it is obvious that whenever such repetitive occurrences are counted, then the sum of positive and negative occurrences must add up to the total number of relevant cases. As far as this objectivist interpretation of probability is concerned, therefore, there is little doubt that classical formalisms provide a suitable conceptual tool to capture the behavior of systems that expresses itself, as experimentally observed, in the form of stable frequency values.

Probabilities, viewed from the perspective of our possible-worlds model, may be considered as the basis of methods providing answers to a question that is related to but different from the undecidable issue of the validity of a hypothesis. Unable to state, because of lack of information, that h is either true or false, we describe instead the behavior of the system in the long run, by calculating the frequency of occurrence under similar circumstances.

Probabilistic reasoning schemes may be generally described as concerned with the computation of the joint probability distribution of several system variables, based on knowledge of the values of related marginal and conditional probability distributions. Whenever the required values are available it is possible, conceptually at least, to derive the required joint distributions. In fact, it may be fairly stated that, once it was understood that such derivation should be the goal of probabilistic reasoning systems, the attention of proponents of that methodological perspective has been almost completely directed toward the development of methods to simplify the required knowledge organization and manipulation [27].

Substantial concerns arise, however, regarding what must be done when the needed probability values are not known. In applied science, when unknown systems and phenomena are investigated, experiments are designed and performed to determine the basic laws of system behavior, which are typically expressed through quantitative relationships. If, based on such knowledge, rational courses of action are chosen, the careful scientist is then able to explain and justify his decisions on the basis of a strong epistemological apparatus supported both by empirical observation and by rational deduction. This scheme, which proceeds from information acquisition to decision making, embodies the experimental method of modern science. From such a perspective, probabilistic laws describe certain aspects of system behavior described by parameters that are estimated using the same methods that are universally accepted and employed in applied science.

Another view of probability, however, regards probability values as expressions of the degree of belief of rational decision makers regarding the validity of hypotheses. This degree of belief is quantified by the amount of money that a rational gambler is willing to bet in a gamble where the payoff, if the unknown truth value turns out to be **true**, is \$1. The probabilistic behavior of these degrees of belief is justified by a number of axiomatic systems [6,35] providing formal support not only to this subjectivist interpretation of probability but also to a decision-making methodology based on the maximization of expected utility. Related axiomatic formulations have been also developed to support the contention that the only correct procedure for updating such beliefs is the Bayes-Laplace rule [5]:

$$\mathbf{Prob}(q|p) = \frac{\mathbf{Prob}(p|q)\,\mathbf{Prob}(q)}{\mathbf{Prob}(p)}.$$

A number of researchers have questioned, in the past, the purportedly rational nature of these axiomatic systems. Their misgivings, which we share, arise both from questions about the rationality of some specific axioms, as noted by Suppes [42], and from observation of the behavior of rational decision-makers(including developers of the axiomatic formalisms) that contradicts the sure-thing principle, as observed by by Allais [1] and Ellsberg [11]. Kyburg [21] has also raised substantial concerns about

the epistemological status and soundness of the subjectivist approach. The axiomatic system of Cox has also been criticized for its assumption that beliefs are measured by a single number [10] and, again, for the less-than-natural character of some axioms [38].

Proponents of this stringent orthodoxy have often argued that behavior departing from their theoretical requirements, however prevalent, is actually irrational. Such a claim, however, suffers from a fundamental methodological flaw. Rationality should be defined in terms of basic requirements that demand proper consideration of two fundamental factors: observed empirical evidence and the laws of logic. By requiring compliance with certain basic tenets of rational behavior, such as the famous avoidance of "dutch books," subjectivist schemes certainly attempt to meet one of these requirements, albeit in a limited fashion, as pointed out by Kyburg [21]. By defining rational behavior as that which results from utilization of the proponent's favorite scheme, the characterization of rationality is subjected to a curious argument that inverts the identity of what is rational with what must be done to ensure rational behavior. This inversion effectively ensures that the expected utility approach would always be considered to be rational: in fact, if any other behavior is observed, it would be, by definition, irrational.

This inversion of premises and conclusions is also apparent in other arguments, based on pragmatic necessity considerations, for the superiority of the subjectivist approach. If decisions, even those to obtain more information, must be made, then the elements required to make the decision (i.e., utility functions and degrees of belief) must be assessed. Conversely, any decision implies that such values have been, whether knowingly or not, chosen in some form or fashion. As a result of this close relation between the assessment of situations and the selection of suitable courses of action, guaranteed by the fact that values of expected utilities (i.e., numbers) may always be totally ordered, it is claimed that the subjectivist approach is the only one among approximate reasoning methods that has a rational decision-theoretic apparatus.

As appealing as such claims may be to some decision-makers, we must note again a curious exchange of roles in the scientific discovery process: decisions no longer follow from empirical observation and rational cogitation; rather, parameters that describe knowledge follow from a practical need to choose suitable actions. However pressing may be the need to derive decisions it should be clear that, in the absence of information, it is usually impossible to determine what is the best course of action. Any randomizing device would, under such circumstances, provide a total ordering of possible choices but there is very little to assure us that any behavior based on such arbitrary basis ought to be called rational.

The ultimate goal of an intelligent system is to take actions based on knowledge about the actual rather than the believed behavior of a real world system. It is

difficult to see why, as noted by Kyburg [22], the latter should be given much attention outside psychological research. If applied science is, as generally admitted, a rational enterprise that seeks to uncover the secrets of the universe and to provide guidelines to take actions based on such knowledge, then it is clearly desirable that intelligent agents, in their quest for similar objectives, follow as closely as possible the essential procedures of the scientific method. The ability to produce decisions regardless of the extent and pertinence of available knowledge should be regarded as a handicap rather than as an advantage of a procedure: a fact readily noticed by those engaged in the solution of important real life problems [12]. As we pointed out before, whenever such knowledge is acquired, it is typically reported using a format that emphasizes the quality of the observational method and the strength of the arguments leading from empirical data to the author's conclusions rather than on the basis of personal confidence expressed by willingness to take gambling risks.

I have made a rather long exposition about the dichotomy between subjectivist and objectivist approaches to probability primarily because I believe this to be a major cause of a controversy that, beyond considerations that are solely germane to probabilistic reasoning, extends to the need for techniques that are not directly based on subjectivist orthodoxy. I have also been motivated by the desire to clearly expose a personal position that is shared by many in the approximate reasoning community but that is also often misleadingly described as being antiprobabilistic.

Far from being antagonistic to one approach for the simple sake of promoting others, my eclectic view is the direct result of practical experience with the development of models of complex systems, and of close familiarity with the application of mathematics to technological problems. Probability is indeed a powerful tool to describe chance-related aspects of the behavior of real-world systems. Recent contributions of probabilists and decision scientists, within and without the context of AI, such as the development of network-oriented procedures for probabilistic reasoning [27], are most important additions to our methodological arsenal.

There are, however, limitations on the capabilities of any tool, whether for system analysis or for any other purpose. As is true of any tool, including all methodologies described in this note, the applicability of probability is limited by its inability to perform functions that lie outside its scope, and by practical constraints on our ability to use it in specific situations. In spite of its unquestionable utility, other approaches also play a significant role in the description of the possible state of affairs. These techniques must not be considered to be competitors of probability but, rather, complementary techniques to enhance the understanding of the real world.

4.2 Generalized Probabilistic Reasoning

Those who worry about the potential lack of applicability of techniques based on conventional probability formalisms do not question the conceptual validity of probability as the appropriate tool to measure the frequency of occurrence of diverse events under various conditions or, in some cases, the strength of belief of decision-makers. Concerns about the problems caused by ignorance of probability values, however, have been expressed continuously since the nineteenth century by such prominent logicians as George Boole [3], and have led to the development of approaches to represent probabilistic ignorance by using subsets of possible probability values.

If, for example, the probability of validity of a proposition p is unknown, an interval probability method will represent such ignorance by assigning the interval $[0, 1]$ as the value of the missing probability. If it is known, on the other hand, that an event has better than even chances of occurring, such knowledge will be represented by the $[0.5, 1]$ interval. More generally, probabilistic knowledge may be represented as a set of possible probability values in a hyperdimensional cube, as in the convex probabilities approach of Kyburg [20].

The corresponding probabilistic calculi are straightforward conceptual extensions of the classic, number based calculus. Such extensions produce, for example, intervals of expected utility values on the basis of knowledge expressed as set of possible probability values. These intervals may be used, in many instances, to rank decisions in the same way that such choices are ordered with number-based schemes. When this ordering is not possible (e.g., overlapping intervals show that under certain scenarios A is preferrable to B, while, in other situations, B is to be preferred), the lack of a clear choice does not imply that the decision-theoretic apparatus is defective. Rather, the methodology is rich enough to tell us precisely how far empirical knowledge, combined with the laws of rational thought, can take us. If, beyond that point, it is imperative to do something—a rather unfortunate set of events—any selection scheme, from that point on, will be as rational as any other (i.e., very little).

Although the manipulation of intervals and sets of possible probability values alleviates some conceptual worries, it hardly helps in terms of the ability to perform the required computations. The situation, unfortunately, is made worse by the need to represent and manipulate probability bounds for subsets without the simplifying help that additivity provides for actual probability values. This unfortunate state of affairs is the primary reason for the popularity that an approach—capable of being interpreted in terms of interval probabilities— enjoys today as one of the major methodologies of approximate reasoning. This approach is the Dempster-Shafer calculus of evidence.

Originally developed by Dempster [7] in the context of statistical studies, the ap-

proach was further developed by Shafer [36] as a non-Bayesian alternative to the representation and manipulation of degrees of belief. Recently [32], application of possible-world semantic models to the interpretation of its major structures has shown that the approach is fully consistent with the classical calculus of probability, including the Bayes-Laplace formula. Smets [40] has also recently reviewed the structures of the calculus of evidence proposing, in addition, unconventional extensions based on a nonprobabilistic concept of belief.

The calculus of evidence may be readily understood using our basic model if it is recalled that, whenever assessing the validity of a hypothesis on the basis of emprical knowledge, there are three possible logical outcomes of any reasoning process: the hypothesis may be proved to be true, the hypothesis may be proved to be false, or the information may be insufficient to make either of those conclusions.

If the notation $\mathbf{K}p$ is used to denote the set of situations, i.e., possible worlds, where p can be proved true, if $\mathbf{K}\neg p$ correspondingly denotes those cases where it can be proved false, and if $\mathbf{I}p$ denotes the set of situations where the truth value of p cannot be established without ambiguity, then it is obvious that any probability function $\mathbf{Prob}(\cdot)$ will satisfy the equation

$$\mathbf{Prob}(\mathbf{K}p) + \mathbf{Prob}(\mathbf{K}\neg p) + \mathbf{Prob}(\mathbf{I}p) = 1\,.$$

Furthermore, since the probability of $\mathbf{I}p$ may be positive, it will be true, in general, that

$$\mathbf{Prob}(\mathbf{K}p) + \mathbf{Prob}(\mathbf{K}\neg p) \leq 1\,.$$

The calculus of evidence is based on the representation of the probabilistic information conveyed by evidence by means of *belief functions*. These functions may be readily interpreted in terms of the above probabilities of provability through the equation

$$\mathbf{Bel}(p) = \mathbf{Prob}(\mathbf{K}p)\,.$$

More importantly, these belief functions are usually expressible in a compact form by means of *basic probability assignments* or *mass functions*. These functions m, which are also defined over propositions, are related to belief functions by the equation

$$\mathbf{Bel}(p) = \sum_{q \Rightarrow p} m(q)\,.$$

The ability to represent and manipulate probability intervals by means of mass functions is the major reason for the appeal of the Dempster-Shafer methodology.

Although, in a typical decision problem, we are interested in the truth of p rather than its provability, lack of adequate information precludes determination of the probability of such truth. In general, however, it may be said that

$$\mathbf{Bel}(p) \leq \mathbf{Prob}(p) \leq 1 - \mathbf{Bel}(\neg p)\,.$$

Furthermore, these bounds cannot be improved.

This interpretation of the Dempster-Shafer calculus as concerned with probabilities of provability, as called by Pearl [27], was first formalized by the author using a possible-worlds model based on the use of a modal logic called epistemic logic. The formal system, which is equivalent to the modal system **S5** [19] used by Moore [25] in his pioneer work on the application of modal logic concepts to artificial intelligence problems, is enhanced by consideration of probability distributions over the set of possible worlds. In particular, the unary operator **K** represents the knowledge of a rational agent to prove that a proposition may be known or proved to be true.

The probability of the set of all possible worlds where a proposition p is the most specific proposition that is known to be true, called the *epistemic set*, corresponds to the values of the mass function. In any possible world, this most specific knowledge is the conjunction of all propositions that are known to be true in that possible world.

The semantic model of the Dempster-Shafer theory also validates the so-called Dempster's rule of combination, which permits the combination of belief and mass functions corresponding to evidential observations made under certain conditions of independence. When such conditions are not valid, use of this formula leads, of course, to erroneous results, often, although incorrectly, considered to be an essential handicap of the evidential reasoning approach, rather than a consequence of its misapplication.

From our perspective the only substantial example of such misapplication is that which results from improper use of the Dempster's rule of conditioning, i.e., a particular use of the rule of combination that is valid only under special circumstances, as a substitute for Bayes' rule. Certain methodological limitations of the calculus of evidence, notably the lack of methods to handle with sufficient generality the counterparts of conventional conditional probabilities, are more worrisome, in our opinion, than any distress arising from its misuse or its supposed lack of a decision-making apparatus.

4.3 Possibilistic Reasoning

Our basic semantic model also provides straightforward interpretations [33] for the major concepts and structures of possibility theory [46,9]: an approach to approximate reasoning derived from multivalued logics [31] and the theory of fuzzy sets [45]. The major formal tool that enhances our understanding of such structures is not a probabilistic measure of set size but, rather, a binary measure of proximity or distance, called a *similarity relation*.

Similarity considerations play a major role in human cognitive processes [44]. In-

formally, all such analogical processes are based on the notion that the validity of some propositions in a given situation extends also to other situations where the same basic conditions are prevalent.

In our model of possibilistic structures, the similarity between states of affairs is expressed by a function that assigns a number between 0 and 1 to every pair of possible worlds. The value of that function $S(w, w')$ for a pair of possible worlds quantifies the extent of resemblance between pairs of situations or scenarios, as evaluated from the viewpoint of the particular problem being considered. In a decision-making problem, for example, the decision maker may define such measures to describe the extent by which the consequences of certain decisions resemble desirable goals or objectives.

The highest similarity value, 1, indicates that, from the perspective of the system being studied, both situations are indistinguishable. The lowest value, 0, indicates that knowlege of what is true in one possible world does not help to derive what is true in the other.

Similarity scales are the measurement sticks used to describe the extent by which certain results may be extrapolated from one possible world to another. Unlike probability functions, which correspond to either measurable properties of physical systems or states of belief of rational agents, the similarity relations simply provide a mechanism to describe resemblance between states of affairs.

Similarity relations may also be regarded as generalizations of the modal-logic notion of *accessibility* or *conceivability* [19] by introduction of multiple binary relations R_α between possible worlds (one for each value of α between 0 and 1), defined by

$$R_\alpha(w, w') \text{ if and only if } S(w, w') \geq \alpha\,.$$

These relations also justify the use of a possibilistic terminology that regards propositions as being possible to some degree, thereby generalizing the classical definition of the modal operator for possible truth in a manner similar to that used by Lewis [23] in his treatment of counterfactual statements.

Certain requirements must be imposed to assure that similarity functions truly represent notions of resemblance between possible situations. Similarities between identical scenarios, for example, should have a value of 1, the highest possible value. Furthermore, if two different possible worlds are to be distinguished by means of similarity values, then it also makes sense to require that their similarity be strictly less than 1. It is likewise natural to require that the similarity between two particular scenarios be a symmetric function, i.e., w resembles w' as much as w' resembles w.

Beyond these properties of reflexivity and symmetry, it is also necessary to require that similarities satisfy a generalized form of transivity. If, given three possible worlds w, w' and w'', the worlds w and w' are highly similar while w' and w'' are also highly

similar, it will be unreasonable to say that w and w'' may be highly dissimilar. The value of $S(w, w'')$ must, therefore, be bounded by below by a function of $S(w, w')$ and $S(w', w'')$, as expressed by the condition

$$S(w, w'') \geq S(w, w') \circledast S(w', w''),$$

which uses the binary operation $\circledast$ to denote the required function.

If certain reasonable requirements are imposed upon the function $\circledast$, it is easy to see that this function has the properties of *triangular norms*, which are usually introduced in multivalued logics [43] to relate the truth value of a conjunction $p \wedge q$ to the degrees of truth of p and q. These functions are motivated, in our model, by considerations that are related solely to metric concepts of proximity and resemblance. Important examples of triangular norms are given by the functions

$$a \circledast b = \min(a, b), \qquad a \circledast b = \max(a + b - 1, 0), \qquad \text{and} \qquad a \circledast b = ab,$$

called the *Zadeh*, *Lukasïewicz*, and *product* triangular norms, respectively.

Similarity functions are trivially related by the relation

$$\delta = 1 - S,$$

to functions δ that have the properties of a distance or metric function. In the particular case where $\circledast$ is the triangular norm of Lukasïewicz, then δ is an ordinary metric or distance, which obeys the well-known triangular inequality

$$\delta(w, w'') \leq \delta(w, w') + \delta(w', w'').$$

If $\circledast$ is the Zadeh triangular norm, on the other hand, the transitivity property is equivalent to the stronger *ultrametric* inequality

$$\delta(w, w'') \leq \max(\,\delta(w, w'), \delta(w', w'')\,).$$

The structures introduced by similarity relations may be readily applied to generalize the subset inclusion relations that are the fundamental basis of deductive reasoning. These inclusion relations are typically expressed by conditional propositions of the form " If q, then p," stating that any state of affairs where q is true is such that p is also true. These conditional propositions, which permit the derivation of true propositions from knowledge of the truth of others by means of the rule of modus ponens, may be also stated using similarity structures by saying that any q-world has a p-world (i.e., itself) that is as similar as possible to it.

The ability to characterize proximity between possible worlds using a continuous scale of similarity provides for a more general characterization of the inclusion relations that hold between subsets of possible worlds (i.e., propositions). If the subset

of q-worlds is not included in that of p-worlds, we may, however, use the similarity structure to quantify the amount of stretching required to reach a p-world from any q-world. The *degree of implication* function defined by the expression

$$\mathbf{I}(p \mid q) = \inf_{w' \vdash q} \sup_{w \vdash p} S(w, w'),$$

which is related to the well-known Hausdorff distance, provides such quantification as the size of the topological neighborhood of p that encloses q, as shown in Figure 2.

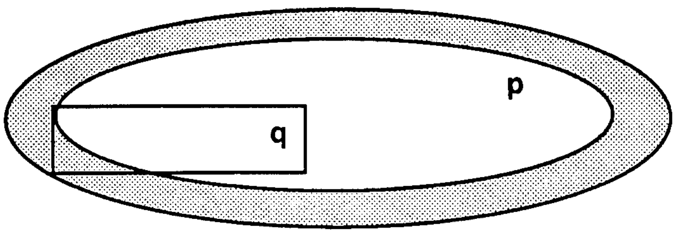


Figure 2: Degree of implication

The ability to express relationships between neighborhoods of different sets of possible worlds or, equivalently, between propositions permits the generalization of the modus ponens by use of the transitive property of the degree of implication function:

$$\mathbf{I}(p \mid r) \geq \mathbf{I}(p \mid q) \circledast \mathbf{I}(q \mid r),$$

illustrated in Figure 3.

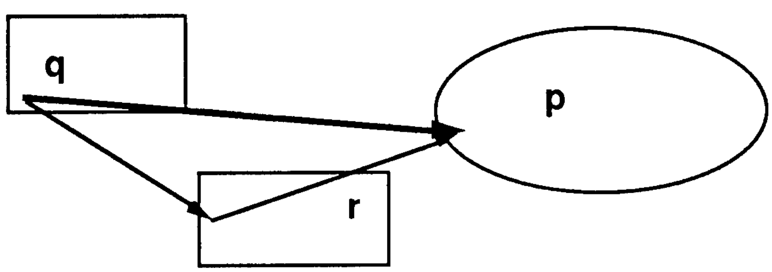


Figure 3: The generalized modus ponens.

The *generalized modus ponens* rule of Zadeh [46] is expressed by means of *possibility distributions*, which are themselves defined in terms of similarities between evidential worlds and those satisfying a given proposition p [33]. From the viewpoint of our similarity-based model, the generalized modus ponens may be thought of as

a sound rule of logical extrapolation that exploits similarities between conceivable scenarios or situations. The fundamental topological structures that permit this type of reasoning are clearly different in character and nature than the measures of set extension that are the conceptual basis of probabilistic reasoning.

In closing, it is important to mention that posibilistic reasoning based on fuzzy logic has led recently to the implementation of a large number of successful commercial products [41]. These systems, which have primarily exploited the applicability of the technology to a variety of control devices, provide a clear indication of the usefulness of these ideas, which now also rest on clearly understandable theoretical foundations.

5 Looking ahead

The ability to explain the role and utility of the major approximate reasoning approaches by use of a unifying framework provides the rational basis to resolve most of the issues about relative importance and necessity. Rather than supporting any partisan contention about the superiority of one methodology over the others, this framework shows instead that a variety of tools are needed to produce effective descriptions of evidence and its implications.

Each methodology may play a significant role in every potential application of approximate reasoning techniques: a role that complements rather than substitutes forf other procedures. In the absence of compelling theoretical arguments for rejecting any approximate reasoning position and in the presence of substantial solid evidence of their usefulness and applicability, it is irrational to maintain positions that are needlesly divisive and polemic.

Recent investigations showing that there exist substantial functional rather than conceptual similarities between the network-oriented methods of conventional probabilistic schemes and the calculus of evidence [37], and indicating that fuzzy-set concepts and multivalued logic may be successfully blended to represent vague knowledge about probabilities [2], clearly point the way toward a more productive research collaboration between approximate reasoning specialists.

This collaboration should stress application of all valid concepts to the solution of practical problems rather than further continuation of the controversy about technological superiority or necessity. In particular, the example set by Japanese researchers in the development of a large number of commercial products of evident applicability illuminates the path that must be followed. The future lies in the solution of practical problems, both because of the direct importance of those problems, and because conceptual developments and clarifications usually follow, as is the case of the work discussed in this note, from the experiences gained producing such solutions. Having

established needed conceptual bases to clarify controversial issues, we hope it is clear that this is the time to apply ideas rather than to continue to argue about them.

References

[1] M. Allais. Le Comportement de l'Homme Rationnel devant le risque: Critique des Postulats et Axiomes de l'Ecole Americaine. (*Behavior of a Rational Man under Risk: Critique of the Postulates and Axioms of the American School.*) Econometrica, 21:503–546, 1953.

[2] P. Bonissone and K. Decker. Selecting uncertainty calculi and granularity: an experiment in trading-off precision and uncertainty. In L.N. Kanaal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, Amsterdam: North Holland, 1986.

[3] G. Boole. *The Laws of Thought*, 1854. New York: Reprinted by Dover Books, 1958.

[4] R. Carnap. *The Logical Foundations of Probability.* Chicago: University of Chicago Press, 1950.

[5] R.T. Cox. On inference and inquiry, an essay in inductive logic. In R.D. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*, Cambridge, Massachussets: MIT Press, 1979.

[6] B. DeFinetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut H. Poincaré*, 7:1–68, 1937.

[7] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Statistics*, 38:325–339, 1967.

[8] R.O. Duda, J. Gaschnig, J., P.E. Hart, K. Konolige, R. Reboh, P. Barrett, and J. Slocum. *Development of a computer-based consultant for mineral exploration.* SRI Project 5821 Final Report and SRI Project 6415 Annual Report. Artificial Intelligence Center, Menlo Park, California: SRI International, 1978.

[9] D. Dubois and H. Prade. An introduction to possibility theory and fuzzy logics. In P. Smets, E.H. Mamdani, D. Dubois, H. Prade, editors, *Non-standard Logics for Automated Reasoning*, 287–326, New York: Academic Press, 1988.

[10] D. Dubois and H. Prade. Comments on P. Garbolino, Probability: a Bayesian viewpoint. In P. Smets, M.R.B. Clarke, C. Froidevaux and J. Campbell, editors, *Positions and Debates in Nonstandard Logics,* New York: Academic Press, forthcoming.

[11] D. Ellsberg. Risk, ambiguity, and the Savage axioms, *The Quarterly Journal of Economics*, 75: 643–669, 1961.

[12] A.R. Feinstein. Clinical Biostatistics XXXIV: The haze of Bayes, the aerial palaces of decisions, and the computerized Ouija board. *Clinical Pharmacology and Therapeutics*, 21:482–496, 1977.

[13] B. Gaines. Fuzzy and probability uncertainty logics. *Inf. Control*, 38:154–169, 1978.

[14] T.D.Garvey, J.D. Lowrance and M.A. Fischler. An inference technique for integrating knowledge from disparate sources. *Proc. 7th. Intern. Joint Conf. on Artificial Intelligence*, Vancouver, British Columbia, 1981.

[15] H. Geffner. *Default Reasoning: Causal and Conditional Theories.* UCLA Cognitive Sciences Laboratory Technical Report (R-137), December 1989 (Ph.D. Dissertation, University of California, Los Angeles, 1989).

[16] I. Hacking. *The Emergence of Probability.* Cambridge, England: Cambridge University Press, 1975.

[17] P.R. Halmos. *Measure Theory.* New York: Springer-Verlag, 1974.

[18] D. Heckerman. Probabilistic interpretations for MYCIN certainty factors. In L.N. Kanaal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, Amsterdam: North Holland, 1986.

[19] G.E. Hughes and M.J. Creswell. *An Introduction to Modal Logic.* New York: Methuen, 1972.

[20] H.E. Kyburg. *Logical Foundations of Statistical Inference.* Dordrecht: Reidel, 1974.

[21] H.E. Kyburg. Subjective probability: criticisms, reflections, and problems. *Journal of Philosophical Logic*, 7:157-180, 1978. Reprinted, with modifications, in *Epistemology and Inference*, Minneapolis: University of Minnesota Press, 1983.

[22] H.E. Kyburg. Interval valued probabilities. In P. Smets, M.R.B. Clarke, C. Froidevaux and J. Campbell, editors, *Positions and Debates in Nonstandard Logics,* New York: Academic Press, forthcoming.

[23] D. Lewis. *Counterfactuals.* Cambridge, Massachusetts: Harvard University Press, 1973.

[24] J. McCarthy. Circumscription: A non-monotonic inference rule. *Artificial Intelligence*, 13:27–40,1980.

[25] R. Moore. *Reasoning about Knowledge and Action.* Technical Note 191, Menlo Park, California: SRI International, 1980.

[26] N.J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1987.

[27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, California: Morgan Kaufmann, 1988.

[28] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: a survey. *Proc. 1st. Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR'89),* Toronto, May 1989, 509–516, San Mateo, California: Morgan Kaufmann, 1989.

[29] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[30] R. Reiter. On close world data bases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, 119–140, New York: Plenum, 1978.

[31] N. Rescher. *Many Valued Logic.* New York: McGraw-Hill, 1969.

[32] E.H. Ruspini. *The Logical Foundations of Evidential Reasoning.* Technical Note No. 408, Artificial Intelligence Center, Menlo Park, California: SRI International, 1987.

[33] E.H. Ruspini. *On the Semantics of Fuzzy Logic.* Technical Note No. 475, Artificial Intelligence Center, Menlo Park, California: SRI International, Menlo Park,1989. *International Journal of Approximate Reasoning,* forthcoming.

[34] W.C. Salmon. Confirmation and relevance, induction, probability and confirmation. In G. Maxwell and R.M. Anderson, editors, *Minnesota Studies in the Philosophy of Science*, volume 6, Minnneapolis: University of Minnesota Press, 1975.

[35] L.J. Savage. *The Foundations of Statistics.* New York: Dover, second revised edition, 1972.

[36] G. Shafer. *A Mathematical Theory of Evidence.* Princeton, New Jersey: Princeton University Press, 1976.

[37] P.P. Shenoy and G. Shafer. An axiomatic framework for Bayesian and belief function propagation. *Proc. 4th. Workshop Uncertainty in Artificial Intelligence*, Minneapolis, Minnesota, 1988.

[38] G. Shafer. Comments on "An inquiry into computer understanding" by Peter Cheeseman. *Computational Intelligence*, 4:121–124, 1988.

[39] E.H. Shortliffe and B. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23: 351–379, 1975.

[40] P. Smets. Belief functions. In P. Smets, M.R.B. Clarke, C. Froidevaux and J. Campbell, editors, *Positions and Debates in Nonstandard Logics,* New York: Academic Press, forthcoming.

[41] M. Sugeno. *Industrial Applications of Fuzzy Control.* Amsterdam: North Holland, 1985.

[42] P. Suppes. The measurement of belief. *J. Roy. Stat. Soc. Ser. B.*, 36:160–175, 1974.

[43] E. Trillas and L. Valverde. On mode and implication in approximate reasoning. In M.M. Gupta, A. Kandel, W. Bandler, J.B. Kiszka, editors, *Approximate Reasoning and Expert Systems*, Amsterdam: North Holland, 157–166, 1985.

[44] A. Tversky. Features of similarity. *Psychological Review*, 84:433–460, 1977.

[45] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8: 338–353, 1965.

[46] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

Understanding Evidential Reasoning

Understanding Evidential Reasoning

Enrique H. Ruspini, John D. Lowrance, Thomas M. Strat

Artificial Intelligence Center
SRI International
Menlo Park, California

ABSTRACT

We address recent criticisms of evidential reasoning, an approach to the analysis of imprecise and uncertain information that is based on the Dempster–Shafer calculus of evidence.

We show that evidential reasoning can be interpreted in terms of classical probability theory and that the Dempster–Shafer calculus of evidence may be considered to be a form of generalized probabilistic reasoning based on the representation of probabilistic ignorance by intervals of possible values. In particular, we emphasize that it is not necessary to resort to nonprobabilistic or subjectivist explanations to justify the validity of the approach.

We answer conceptual criticisms of evidential reasoning primarily on the basis of the criticism's confusion between the current state of development of the theory – mainly theoretical limitations in the treatment of conditional information – and its potential usefulness in treating a wide variety of uncertainty analysis problems. Similarly, we indicate that the supposed lack of decision-support schemes of generalized probability approaches is not a theoretical handicap but rather an indication of basic informational shortcomings that is a desirable asset of any formal approximate reasoning approach. We also point to potential shortcomings of the underlying representation scheme to treat general probabilistic reasoning problems.

We also consider methodological criticisms of the approach, focusing primarily on the alleged counterintuitive nature of Dempster's combination formula, showing that such results are the result of its misapplication. We also address issues of complexity and validity of scope of the calculus of evidence.

Address correspondence to E. H. Ruspini, Artificial Intelligence Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025-3493.

This work was supported in part by the Defense Advanced Research Project Agency under Contract N00039-88-C-0248, in part by the Air Force Office of Scientific Research under Contract No. F49620-89-K-0001, and in part by the United States Army Research Office under Contract No. DAAL03-89-K-0156.

International Journal of Approximate Reasoning 1992; 6:401–424

655 Avenue of the Americas, New York, NY 10010 0888-613X/92/$05.00

KEYWORDS: *evidential reasoning, probabilistic reasoning, belief functions, Dempster-Shafer theory, calculus of evidence*

1. INTRODUCTION

If artificially intelligent systems are to produce adequate assessments of the state and behavior of the real world, they must cope with information and knowledge that is characterized by varying degrees of uncertainty, ignorance, and correctness. To address this need, we have developed a technology called *evidential reasoning*. It is formally based upon the Dempster–Shafer [1] theory of belief functions, it has been implemented as a domain-independent automated reasoning system, and it has been successfully applied to a range of real-world problems (Lowrance *et al.* [2]). Yet, its reliance on belief functions has drawn criticism.

Our choice of an approach based on the Dempster–Shafer theory was not arbitrary. We believe that theory confers important methodological advantages, such as its ability to represent ignorance in a direct and straightforward fashion, its consistency with classical probability theory, its compatibility with Boolean logic, and its manageable computational complexity. At the same time, we recognize that other approaches may also complement and augment the assessments provided by evidential reasoning.

We examine several criticisms of belief functions that have appeared in the literature, discussing first the fundamental theoretical bases supporting the belief function approach and justifying its use in terms of the requirements imposed by ignorance of certain probability distributions. We consider the nature of Dempster's rule of combination and argue that negative assessments either misinterpret the nature of the distributions being combined or ignore the basic independence assumptions that ensure its validity. We stress also that it is not necessary to rely on explanations that are either nonprobabilistic or subjective to justify the validity of the Dempster–Shafer calculus of evidence.

Furthermore, we show that certain apparently counterintuitive properties of the approach (e.g., the "spoiled sandwich" paradox) are the natural consequence of considering families of possible probability distributions that solve an approximate reasoning problem. In the context of this discussion, we indicate also the inherent pitfalls of "axiomatic" approaches that accept or reject methodologies on the basis of their compliance with allegedly intuitive principles.

We also answer critiques based on the computational complexity of the belief function approach. Such criticisms claim that the complexity of probabilistic knowledge representations grows exponentially with the size of the frame, thus making the theory unsuited for automated reasoning. Other comments addressed in our presentation center on limitations on the representational ability of belief functions and the lack of certain methodological capabilities (e.g., decision-making mechanisms).

Despite the criticism that belief functions have drawn, we believe that evidential reasoning is well founded and that it may be effectively applied to the solution of a broad range of important practical problems.

Most of our comments will be made in direct reply Pearl's recent criticism of the belief function approach [3], because we feel that his paper encompasses most of the major worries and concerns expressed about the calculus of evidence. Although most of the discussion in this paper consists of direct responses to issues raised by Pearl and others, our overall objective is considerably broader. Our answers are motivated by the remarks of DeGroot, quoted by Pearl at the conclusion of his work, about the need to use our methodological approaches "with the utmost care and in accordance with the highest ethical standards." Our aim, like Pearl's, is to enlighten and clarify, through careful discussion of rather subtle and delicate issues, rather than to engage in dogmatic defense of one approach to the detriment of another. It is our earnest hope that this work, in conjunction with other evaluations of the belief function approach, will lead to a better understanding of its foundations, capabilities, and limitations.

2. ON THEORETICAL SOUNDNESS

The theory of belief functions was originated by Dempster [4] in the context of statistical research. The use of the term "belief," together with its subjectivist connotations, is due to Shafer [1], who first applied the theory to the analysis of imprecise and uncertain evidence.

Although much skepticism has been voiced about the naturality of belief functions and their agreement with conventional probabilistic approaches, its theoretical bases are provided by a simple consideration of the role of evidence as a basic information carrier.

In classical probabilistic treatments, it is assumed that, under certain evidential conditions $\mathscr{E}$,[1] the value $P(p \mid \mathscr{E})$ of the likelihood of a particular statement p is known. This view of evidence, adequate to represent the informational conditions of most controlled experimental setups, fails to adequately model the effects that acquiring similar information has on our state of knowledge when the state of the world cannot be so readily manipulated.

In such circumstances, whenever the evidence $\mathscr{E}$ is observed, three possible informational outcomes may result from examination of further information

[1] Throughout this paper, the symbol $\mathscr{E}$ is used to denote available evidence, that is, a collection of propositions about the real world that are known to be true either as a result of direct observation or as the consequences of applicable background knowledge.

that later turns out to improve our state of knowledge: Either p is found to be true, $\neg p$ is found to be true (i.e., p is false), or such information is insufficient to determine the truth value of p. Use of modal logic concepts, which are the bases of the formal model of Ruspini [5] suggests the use of the notation $\mathbf{K}p$, $\mathbf{K}\neg p$, and $\mathbf{I}p$ to identify these outcomes. Since these alternatives are exclusive, it is clear that

$$P(\mathbf{K}p) + P(\mathbf{K}\neg p) + P(\mathbf{I}p) = 1.$$

Furthermore, since the probability of $\mathbf{I}p$ may be positive, it will be true, in general, that

$$P(\mathbf{K}p) + P(\mathbf{K}\neg p) \leq 1.$$

This model, based on a combination of classical probability methods and the modal logic *S5* (Hughes and Cresswell [6], Moore [7]), essentially provides—through the logical notation of possible world—a meaning for the unary operator **K** as the representation of the state of knowledge of a statistician who is estimating the probability of truth of diverse propositions $\{p, q, \cdots\}$ under evidential conditions.

This statistician estimates those distributions by considering multiple samples of the state or behavior of a real-world system. Using, for each sample, additional information collected through further experimentation, the statistician may then establish or not the validity of a proposition p. If he is rather lucky, our statistician will find himself in the ideal situation where he can actually "know"[2] or "prove" that the real world is in a state s that is described to the best level of detail that is necessary to understand its behavior (i.e., a "possible world"). This is the state of knowledge usually attained, under perfect laboratory conditions, when experimental samples are fully analyzed and when the outcome of such analyses is classified in terms of a set of exhaustive and mutually exclusive alternatives.

Under less desirable epistemological circumstances, however, the statistician will only be able to prove that a less specific proposition q is true. In the extreme case where no further information exists, he will be forced to say that his knowledge is limited to that provided by the evidence $\mathscr{E}$, or that it is "vacuous."

All samples so analyzed, however, can be classified as to the "most specific knowledge" that could be determined in each case. The corresponding probability measure of the set $e(p)$ of samples where the proposition p was the most specific knowledge (called an *epistemic set* by Ruspini) corresponds, in

[2] Note that, in the context of epistemic logics such as *S5*, the operator K behaves as a logical necessity operator. "Knowing" a proposition simply means that observations logically imply such a proposition or that it is *necessarily* true.

Shafer's framework, to the value $m(p)$ of a mass function m, that is,

$$m(p) = P(e(p)).$$

Correspondingly, the probability that p was "known" to be true during statistical experimentation corresponds to the value Bel(p) of Shafer's belief functions,

$$Bel(p) = P(\mathbf{K}p).$$

The connection between the ability of our statistician to know that p was true and the belief and mass functions that he estimates through experimentation justifies both the expression *epistemic probability* introduced by Ruspini [5] to describe the underlying probabilities defined over a particular set of situations or scenarios $\mathbf{K}p$ (called the *epistemic universe*) and the description of the functions as being "probabilities of provability" or "probabilities of necessity" by Pearl [8], following a suggestion by Fagin and Halpern [9].

In short, all such interpretations are equivalent to the original model of Ruspini, where a rational agent was able to prove the truth of different propositions under different information circumstances that were found to prevail, during his statistical experiment, with different frequencies of occurrence.[3]

Since the ability to prove a proposition q entails the ability to prove any proposition p that is implied by q, it should be clear that

$$Bel(p) = \sum_{q \Rightarrow p} m(q),$$

which is the fundamental equation relating the basic structures of the calculus of evidence. It is also true that

$$Bel(p) \leq P(p) \leq 1 - Bel(\neg p),$$

providing bounds for the probability of p that may not be improved. This ability to manipulate probability intervals by means of the compact representation scheme of mass functions is the major reason for the appeal of the Dempster–Shafer methodology.

[3] Note, however, that while use of the terms "knowability," "provability," and "necessity" does much to provide adequate semantics to the calculus of evidence, their loose usage leads to unnecessary confusion. For example, in his recent criticism [3], Pearl takes some questionable semantic license with the term "necessity," mentioning, for example, the probability that a decision "will have to be made out of compelling necessity." Such "pragmatic" necessity does not have anything to do, of course, with the "logical necessity" that underlies the Dempster–Shafer theory, that is, the necessary truth of a proposition given available evidence.

While the above discussion clarifies the nature of the statistician's knowledge modeled by belief and mass functions, doubts might still remain as to their utility to those who were not involved in their statistical estimation process. Such usage is, however, that made of any other probabilistic information. The analyst who observes $\mathscr{E}$ does not have the luxury that was available to the statistician estimating epistemic probabilities, that is, the ability to collect additional information that permits a more detailed characterization of the state of the world, for the same reasons that the user of statistical tables is unable to utilize the raw data of the estimating statistician. Under such circumstances, the analyst is forced to rely on the probabilistic estimates provided by the statistician, which are believed on the basis of the assumed regularity of the repetitive behavior of the system: the epistemological cornerstone of probabilistic reasoning.

In other words, the "probability of provability" is the best information that is available to the analyst; an observation that disposes not only of questions about its role in probabilistic reasoning, but also of Pearl's worries about its use in lieu of the obviously more desirable "probability of truth" [3]:

> why we should concern ourselves with the probability that the evidence implies A, rather than the probability that A is true, given the evidence.

Clearly, we would prefer having the latter, but unfortunately, we can measure only the former.

Our interpretation of the major evidential functions and structures also quickly disposes of erroneous arguments based on unintended interpretations of the intervals defined by belief functions. Each such interval represents ignorance of a single probability value for a proposition p under fixed evidential conditions $\mathscr{E}$. If critics choose, for example, to interpret such intervals as the possible values that conditional probabilities might attain when further evidence is collected, as suggested by Pearl [10], belief functions will not, indeed, behave according to such unintended semantics.

In closing this section, it is important to mention other alternative views of the structures of the calculus of evidence such as that recently proposed by Smets [11], which are based on a nonprobabilistic concept of belief. Although those models are interesting on the strength of their own virtues, we still emphasize that such interpretations are not required to reconcile the calculus of evidence with conventional probability theory.

In consideration of our ability to reconcile all structures and formulas of the calculus of evidence, including Dempster's formula, with conventional probability structures, such as inner and outer probabilities, we do not feel strongly compelled to accept alternative epistemic interpretations. Our skepticism in this regard is further supported by the observation that, often, such epistemological alternatives are the result of misunderstandings about the role of certain

evidential formulas and processes (e.g., normalization). For the same reasons, we remain unconvinced about the need to assign alternative interpretations to the structures of calculus of evidence or to its functions, as is the recent suggestion of Halpern and Fagin [12], which is echoed by Pearl [3].

3. ON DECISION SUPPORT

A criticism of a more fundamental nature of the calculus of evidence is often raised regarding the output of generalized interval-probability approaches. Since these methods often fail, because of basic knowledge deficiencies, to rank decision choices by the value of some measure that quantifies the desirability of each choice (e.g., expected utility), then it is said that they lack a decision-theoretic apparatus.

Although these arguments correctly point to the basic knowledge requirement that most decision problems entail—if a rational choice is to be made, then we must have a proper informational basis to do it—this obvious consideration is twisted to argue for the necessity to estimate unknown probability and utility values when they are not available. We do not think that this *pragmatic necessity* argument is either sound or compelling.

In our view, the calculus of evidence may be used in a straightforward fashion to produce intervals of possible utility values. When such intervals overlap and cannot be ordered, this fact simply reflects a basic deficiency in our knowledge. We look down upon "pragmatic justifications" with the same concern that any experimental scientist must show about proposals to guess what he has not measured: The ability to make decisions in the absence of knowledge is, in our view, a handicap rather than an advantage of any method.

Far from lacking a decision-theoretic methodology, our approach provides an understandable quantification of the undesirable effects that poor information has on our decision-making ability, ordering decisions whenever it is rationally possible but advising us that such ranking is not possible if our knowledge is insufficient. In brief, our approach not only supports decision making but, through its built-in sensitivity analysis features, helps us to determine what must be done to reach a happier epistemological state.[4]

4. ON DEMPSTER'S RULE OF COMBINATION

The semantic model of the Dempster–Shafer theory also validates the so-called Dempster's rule of combination, which permits the combination of

[4] For an example of an approach that incorporates decision-maker preferences into the framework of the belief function calculus, the reader is referred to a recent paper by Strat [13].

belief and mass functions corresponding to different evidential observations made under certain conditions of independence. When such conditions are not valid, use of this formula leads, of course, to erroneous results, often, although incorrectly, considered to be an essential handicap of the evidential reasoning approach rather than a consequence of its misapplication.

The Dempster formula is, currently, the principal evidence integration mechanism of the belief function approach. It was derived in the context of a basic model of the effect of probabilistic evidence that correctly interprets such evidence as constraints on probability values rather than as the source of the actual values, which are typically undetermined. It may be described as an expression that, under certain conditions of independence, yields bounds for the conditional probability distribution $P(\cdot \mid \mathscr{E}_1, \mathscr{E}_2)$ on the basis of similar bounds for the probability distributions $P(\cdot \mid \mathscr{E}_1)$ and $P(\cdot \mid \mathscr{E}_2)$.

To understand the conceptual bases for Dempster's formula of combination and its consistence with conventional probability, we resort to a generalization of the logical model used before to derive the basic relations of the calculus of evidence. Instead of considering a single epistemic operator, corresponding to a single statistician or observer, we will consider two such rational agents, with their knowledge modeled by means of two operators $\mathbf{K}_1$ and $\mathbf{K}_2$. Each of these rational agents will be assumed to be ignorant of the knowledge possessed by the other, that is, as if they were statisticians performing independent experiments under different evidential conditions $\mathscr{E}_1$ and $\mathscr{E}_2$. Their common knowledge, however, will be modeled by means of a nonindexed operator $\mathbf{K}$ corresponding to a third reliable agent that aggregates the statistical knowledge gathered by the other two.

Clearly, in a given applicable situation (i.e., the first agent observes $\mathscr{E}_1$ and the second agent observes $\mathscr{E}_2$), the integrating agent, who does not add any knowledge of his own, will be able to prove (or to "know" the truth of) a proposition p if the other agents provide individual items of information that, when combined (i.e., conjoined), imply p, as expressed by the basic combination axiom:

$\mathbf{K}p$ is true if and only if there exist sentences p_1 and p_2 such that $\mathbf{K}_1 p_1$ and $\mathbf{K}_2 p_2$ are true, and such that $p_1 \wedge p_2 \Rightarrow p$.

Using our three operators to generate all possible (i.e., logically consistent) states of knowledge that may be attained by each of the three agents while assessing the state of a real system, we may say that each of them has, as was the case before, knowledge about the real world that may be represented by the "most specific"[5] propositions p_1, p_2, and p that each has been able to prove (with p being obviously more specific than either p_1 or p_2). In the terminol-

[5] Note that such most-specific knowledge always exists and is unique except for logical equivalences because the conjunction of all proved theorems is itself a theorem.

ogy of Ruspini's semantic model, each of the agents is in an epistemic state, denoted by $e(p)$, $e_1(p_1)$, and $e_2(p_2)$, respectively, each corresponding to the set of all conceivable states of the real world (i.e., possible worlds) having such knowledge characteristics.

The following important set equation relating all of these types of epistemic sets as subsets of our enhanced epistemic universe is the basis for the derivation of various evidential combination formulas,

$$e(p) = \bigcup_{p_1 \wedge p_2 = p} (e_1(p_1) \cap e_2(p_2))$$

of which the Dempster combination formula,

$$m(p) = \kappa \sum_{p_1 \wedge p_2 = p} m_1(p_1) m_2(p_2),$$

where

$$m(p) = P(e(p) \mid \mathscr{E}_1, \mathscr{E}_2),$$

$$m_1(p_1) = P(e_1(p_1) \mid \mathscr{E}_1), \qquad m_2(p_2) = P(e_2(p_2) \mid \mathscr{E}_2)$$

and where κ is a multiplicative factor, is the best known and used.

Before reviewing the actual process leading to the derivation of Dempster's formula, it is important to pause and reflect upon the nature of the above set-theoretic equation and its usefulness to derive evidence combination formulas.

We may first note that this equation has been derived as a relation between subsets of possible "epistemological states" that is valid regardless of any assumptions about probabilistic structures and their properties (e.g., independence). As such, it provides the bases not only for the derivation of Dempster's formula but actually for a variety of formulas that bound possible probability values within and outside the structures of the Dempster–Shafer theory.

Basically, this formula provides the basis to extend a probability function P that is known over subsets of the form $e_1(p_1)$ and $e_2(p_2)$ (i.e., over two *σ-algebras*), to the set of unions of sets of the form $e_1(p_1) \cap e_2(p_2)$ (i.e., another σ-algebra). If such extension can be made uniquely—as is the case for Dempster's formula—the resulting extension may be used to generate both the conditional probability $P(\cdot \mid \mathscr{E}_1, \mathscr{E}_2)$ and its associated bounds Bel and Pl, which are fully compliant with Shafer's axioms. In other less fortunate cases (e.g., dependent evidence), such extension is not unique, and the lower envelope of the possible extensions, which is not a probability, will lead to bounds that do not satisfy the axioms of the calculus of evidence.

This equation is now being used to extend the evidential calculus approach by generalization of the notion of conditional probability by study of the

probabilistic relations that define dependencies between the different types of epistemic sets [i.e., $e(p)$, $e_1(p_1)$, and $e_2(p_2)$]. Pearl [3], however, believes, apparently as the result of his examination of the role of *compatibility relations* in the calculus of evidence, that this approach is essentially limited in its expressive ability to set-theoretic relations between epistemic sets, which correspond to classical logical conditional statements (i.e., material implications).

In fact, it can be easily seen from our epistemic identity that whenever the conditional probabilities $P(e_2(p_2) \mid e_1(p_1))$ and $P(e_1(p_1) \mid e_2(p_2))$ are restricted to take the values 0 or 1,[6] this identity can be used to map one body of evidence into another, by means of the compatibility relations that such probabilities define.

Since under these assumptions, however, there can be only one proposition p_2 for every proposition p_1 such that $P(e_2(p_2) \mid e_1(p_1)) = 1$, and vice versa, then the compatibility relation that is so defined can be characterized by several implications of the form

$$e_1(p_1) \Rightarrow e_2(p_2)$$

and of the form

$$e_2(q_2) \Rightarrow e_1(q_1)$$

between knowledge states of one observer and knowledge states of the other that are useful to "transfer mass" between propositions. This correspondence must be contrasted with that following from the limited interpretation given by Pearl, who, from knowledge of

$$e_1(p_1) \Rightarrow e_2(p_2),$$

concludes (by contraposition), correctly but narrowly, that

$$\neg e_2(p_2) \Rightarrow \neg e_1(p_1)$$

and proceeds then to attach all material implication paradoxes (e.g., the "ravens paradox") to the calculus of evidence as if they were an essential methodological bane. If that were the case—clearly it is not—the same concerns should be raised about the use of conditionals in conventional probability calculus.

The second observation that can be made about the nature of evidence combination, in general, and the role of our basic set identity to generate combination formulas, in particular, is that while the functions to be combined

[6] It can be shown from the definition of epistemic sets that, under such conditions, knowledge of $P(e_2(p_2) \mid e_1(p_1))$ suffices to derive $P(e_1(p_1) \mid e_2(p_2))$.

are conditional probabilities over two different evidential sets $\mathscr{E}_1$ and $\mathscr{E}_2$ (i.e., the evidence observed by two agents), the desired integrated probability is a distribution over $\mathscr{E}_1 \cap \mathscr{E}_2$ (since we know that both observations are correct). Except for unusual cases, however, computation of $P(\cdot \mid \mathscr{E}_1, \mathscr{E}_2)$ entails a "normalization" operation that is fully consistent with the calculus of probability. Most of the normalization "paradoxes" are the result of misunderstanding about what is being combined: two different conditional probabilities rather than two different lower and upper bounds of the same probability function.[7]

Focusing now on the rationale for Dempster's formula, we should notice first that the epistemic sets $e_1(p_1)$ and $e_2(p_2)$ are such that

$$e_1(p_1) \subseteq \mathscr{E}_1, \qquad e_2(p_2) \subseteq \mathscr{E}_2,$$

that is, the possible knowledge states of each statistician include awareness of the truth of the evidence that is observed by each. Furthermore,

$$\mathscr{E}_1 = \bigcup_{p_1} e_1(p_1), \qquad \mathscr{E}_2 = \bigcup_{p_2} e_2(p_2),$$

where $p_1 \Rightarrow \mathscr{E}_1$ and $p_2 \Rightarrow \mathscr{E}_2$; that is, each statistician knows something that implies that his evidential observation is true (otherwise he would not be "counting" that sample).[8]

Assume now that there exists a probability distribution P defined over the space of all possible epistemic states for our observing statisticians and our "integrating" agent. Each such epistemic state is a possible world that corresponds to a possible state of the world and to a possible state of knowledge for each agent that, in addition, is consistent with the laws of logic. We will assume now that, whenever $p_1 \Rightarrow \mathscr{E}_1$ and $p_2 \Rightarrow \mathscr{E}_2$,

$$P(e_1(p_1) \cap e_2(p_2)) = \begin{cases} P(e_1(p_1))P(e_2(p_2)) & \text{if } p_1 \wedge p_2 \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

This assumption simply states that when $\mathscr{E}_1$ and $\mathscr{E}_2$ are both true, the probability that a rational observer will be in a particular knowledge, or epistemic, state does not provide any information about the probability of the epistemic state of the other agent (i.e., beyond ruling out logical impossibilities). In purely formal terms, we may say that knowledge of values of P over

[7] It is fair to say that much of the skepticism raised by the normalization used in Dempster's formula can be traced to the exposition given by Shafer [1], which suggests a nonprobabilistic method of evidence combination.

[8] Recall that our observers, or rational agents, are statisticians estimating properties of certain statistical distributions by classifying each sample using their evidence and additional sample-dependent knowledge.

sets of the form $e_1(p_1)$ does not provide any indication, beyond exclusion of logical impossibilities, of the values of P over sets of the form $e_2(p_2)$ and vice versa. The epistemic states of our two agents may be said, therefore, to be unrelated in that knowledge of the state of one of our observers (by our integrating agent) does not provide any information about the state of the other, save for elimination of logical impossibilities.

Noting now that

$$P(e_1(p_1) \mid \mathscr{E}_1) = \frac{P(e_1(p_1))}{P(\mathscr{E}_1)}, \qquad P(e_2(p_2) \mid \mathscr{E}_2) = \frac{P(e_2(p_2))}{P(\mathscr{E}_2)},$$

$$P(e_1(p_1) \cap e_2(p_2) \mid \mathscr{E}_1, \mathscr{E}_2) = \frac{P(e_1(p_1) \cap e_2(p_2))}{P(\mathscr{E}_1 \cap \mathscr{E}_2)},$$

then, whenever $p_1 \wedge p_2 \neq \varnothing$,

$$\begin{aligned} P(e_1(p_1) \cap e_2(p_2) \mid \mathscr{E}_1 \mathscr{E}_2) &= \kappa P(e_1(p_1) \mid \mathscr{E}_1) P(e_2(p_2) \mid \mathscr{E}_2) \\ &= \kappa m_1(p_1) m_2(p_2), \end{aligned}$$

from which the Dempster's formula readily follows.

The normalization factor

$$\kappa = \frac{P(\mathscr{E}_1) P(\mathscr{E}_2)}{P(\mathscr{E}_1 \cap \mathscr{E}_2)}$$

has been the object of considerable concern on the part of both skeptics and proponents of the calculus of evidence. The above expression, however, provides the rationale for its use while disposing of arguments about its alleged inconsistence with the probability calculus. In that expression, the denominator $P(\mathscr{E}_1 \cap \mathscr{E}_2)$ appears as the consequence of the need to derive probability distribution estimates with respect to the intersection of the two observed evidences $\mathscr{E}_1$ and $\mathscr{E}_2$. The numerator of that expression simply reflects the need to combine conditional distributions over the same reference set (i.e., the epistemic universe) while our probabilistic knowledge is expressed over two of its subsets (i.e., $\mathscr{E}_1$ and $\mathscr{E}_2$).

The essence of the conditions that lend validity to the Dempster formula may be summarized by saying that the formula's usefulness is confined to the limited, but rather important, cases where estimates of probabilistic likelihood have been formulated by two rational agents on the bases of independent observations while ignoring the evidence available to the other.

If our integrating agent is thought of as being concerned with estimating the probabilities of certain events when both $\mathscr{E}_1$ and $\mathscr{E}_2$ are true, then we may say that, whenever the conditions validating Dempster's formula hold, knowledge

of the fact that a particular sample satisfies p_1 tells the agent nothing about the likelihood of p_2 (unless, of course, p_1 happens to be logically inconsistent with p_2). Furthermore, whenever our integrating agent is done with his job, he should find out that estimating this joint distribution (i.e., over $\mathscr{E}_1 \cap \mathscr{E}_2$) could have been accomplished in an easier fashion by estimating the marginal distributions over $\mathscr{E}_1$ and $\mathscr{E}_2$ and deriving the joint distribution by multiplication and normalization.

Other accounts supporting the validity of Dempster's formula and its consistence with the probability calculus have been advanced by several authors. A particularly compelling justification has recently been given by Wilson [14].

5. ON "PARADOXES"

Criticisms of the Dempster formula may be broadly characterized as being the consequence of basic misunderstandings about either its meaning or its validity.

In this section, we examine three alleged paradoxes of the theory, showing that the purported inconsistencies are actually the results of conceptual misunderstandings or misrepresentations of the positions of those who, while generally supporting the calculus of evidence, are concerned with its possible misapplication.

5.1 The Three Prisoners Problem

Turning our attention first to concerns about the validity of Dempster's formula, we note that, in general, such examples ignore its scope of applicability, producing counterintuitive results that are then used to dismiss the methodology as inadequate. Among those, the three prisoners problem discussed by Diaconis and Zabell [15] has been perhaps the most quoted and discussed.

This problem is one of a variety of examples in which the combination formula is used as a *conditioning formula* by assuming that one of the mass distributions being combined simply assigns all of its mass to a proposition p in the frame of discernment. Combination of such a simple support function with another mass function associated with a belief function Bel(·) leads to the conditioning formula

$$Bel(q \mid p) = \frac{Bel(q \vee \neg p) - Bel(\neg p)}{1 - Bel(\neg p)}$$

In the particular case of the three prisoners problem, which is concerned with the guilt or innocence of a prisoner who has been chosen (by the warden) as the guilty party by random draw among three candidates A_1, A_2, and A_3,

our "logical space" or frame of discernment is simply the Boolean algebra induced by the three noncompatible propositions

Prisoner A_i has been found guilty

where $i = 1, 2, 3$. Since only one of the three prisoners is chosen by the warden, we clearly have

$$P(p_i) = 1/3, \qquad i = 1, 2, 3.$$

(Note that P is actually a classical, additive, probability distribution.)

Prisoner A_1 now asks the jailer to name one of the innocent prisoners (other than A_1), arguing that such information would clearly be of little help to him as an indicator of his potential fate. As Pearl notes, if q stands for the proposition "The jailer names A_2 as one of the innocent," then application of the conditioning rule leads to the result

$$Bel(p_1 \mid q) = Pl(p_1 \mid q) = 1/2$$

indicating that the conditional probability $P(p_1 \mid q)$ must be exactly 1/2, instead of the "correct solution,"

$$0 \le P(p_1 \mid q) \le 1/2,$$

while also saying, against the correct intuition of A_1, that his chances of guilt have been increased as the result of the irrelevant information provided by the jailer. From such an observation, Pearl concludes that the formula is seriously flawed, both because of the counterintuitive result that it produces and for its "collapsing" of a family of solutions into a single value.

Before proceeding to the discussion of Pearl's concerns, we may note, in passing, that this problem has been well known as a source of paradoxes and incorrect solutions within the scope of the conventional probability calculus (Bar-Hillel and Falk [16]) quite independently of any issues of validity of its treatment using the Dempster–Shafer calculus. The explanations given to describe the conceptual errors leading to incorrect classical treatments resemble to some extent those that shed light on the inapplicability of Dempster's formula.

Returning now to the role of Dempster's formula in this problem, we first observe that although, at first glance, the distributions representing the jailer's and warden's choices seem independent, it is actually impossible for the jailer to tell A_1 that A_2 is one of those to be spared if all he knows is that the Warden is choosing the guilty party by random draw (i.e., he needs to know exactly who is the one chosen for punishment). To use the terminology of Ruspini's model, the probability of A_2 being named as one of the innocent depends on the epistemic state of the warden, thus violating the independence

assumptions of Dempster's formula. If all possible combinations of truth values for the propositions p_i, $i = 1, 2, 3$, and q are tabulated, together with their probabilities, as is done in Table 1, then it is clear that

$$P(q \mid p_3) = 1, \qquad P(q) = (1/3)(1 + \alpha),$$

where $0 \leq \alpha \leq 1$ represents the unknown probability that the jailer will choose to name A_2 rather than A_3 as innocent if A_1 is actually the one chosen by the warden as guilty.

But then,

$$P(q \mid p_3) \neq P(q)$$

violating the assumptions, discussed above, that validate the use of Dempster's formula [i.e., $P(e_2(p_2) \mid e_1(p_1)) \neq P(e_2(p_2))$]. There is not, therefore, "total mystery," as Pearl says, as to the incorrect results obtained using Dempster's formula. Because it fails to be applicable, there should be little wonder that it leads to an apparent paradox.

Although, as clearly shown by this discussion, the incorrect treatment of the three prisoners problem fails to invalidate Dempster's rule of combination, we share the concern of Pearl and others about its wide misapplication, particularly when it is used indiscriminately to generate conditional distributions. In our research, we are endeavoring to extend the original theory to produce expressions to produce and utilize conditional belief information (Ruspini [17]) that incorporates known dependencies between evidential bodies. These formulas are intended to provide better interval estimates than the typically uninformative bounds that are supplied by strict derivation of bounds in the absence of additional information by the expression

$$Bel(q \mid p) = \frac{Bel(p \wedge q)}{Bel(p \wedge q) + Pl(p \wedge \neg q)},$$

which is mentioned in Dempster's original paper [4] and that has been the object of recent concern by several authors (de Campos et al. [18], Halpern and Fagin [12]).

In closing, we believe it is important to address other concerns of Pearl, apparently going beyond the three prisoners problem, about the counterintu-

Table 1. Possible Worlds in the Three Prisoners Problem

Possible World	Warden's Choice	Jailer Identifies	Probability
W_1	A_1	A_2	$(1/3)\alpha$
W_2	A_1	A_3	$(1/3)(1-\alpha)$
W_3	A_2	A_3	$1/3$
W_4	A_3	A_2	$1/3$

itive nature of the "collapse" that usage of the Dempster formula often yields, which is manifested by production of a single conditional probability distribution when conditioning multiple members of a family $\mathscr{P}$ of probabilities over some specific subset q. Just as it is true that all members of the family of distributions

$$\mathscr{P} = \{P_t : t \text{ in } [0, 1]\},$$

defined in the set $X = \{a, b, c\}$ by the expression

$$P_t(x) = \begin{cases} 1/2t, & \text{if } x = a, \\ 1/2(1-t), & \text{if } x = b, \\ 1/2, & \text{if } x = c, \end{cases}$$

are such that $P_t(\{a, b\}) = 1/2$, despite their variability over other subsets, it is also true that an extensive family of distributions may collapse into a single conditional probability without violating any rational or probabilistic principles. Such "invariants" are, in fact, desirable as elements that simplify the analysis of an otherwise complex probabilistic problem. For these reasons, we believe that if Dempster's conditioning formula is applicable, its reduction of the variability of probability values should not be a particular cause for concern as to its validity.

5.2. The Spoiled Sandwich

While discussing the suitability of the calculus of evidence either as a form of generalized probabilistic calculus or as a new theory that intends to capture a novel notion of belief, Pearl [3] again faults the approach for failing to satisfy the following rationality principle originally stated by Aleilunas [19]:

> If two diametrically opposed assumptions yield two different degrees of belief in a proposition Q, then the unconditional degree of belief merited by Q should be somewhere between the two.

As natural as such a principle might look at first, the following simple and clever example from Wilson [20] clearly shows that it is neither intuitive nor appealing but points instead to the pitfalls of creating or supporting one's favorite scheme on the strength of supposedly rational axioms.

Let $X = \{a, b, c, d\}$ with $A = \{a, b\}$ and $B = \{a, c\}$, so that $\overline{B} = \{b, d\}$. Consider the family of probability distributions in X,

$$\mathscr{P} = \{P_t : t \text{ in } [0, 1]\},$$

indexed by a parameter t in [0, 1] and defined by

$$P_t(\{a\}) = 1/2t,$$

$$P_t(\{b\}) = 1/2(1-t),$$

$$P_t(\{c\}) = 1/4,$$

$$P_t(\{d\}) = 1/4,$$

and let

$$P^* = \inf_t \{P_t\}.$$

Then, clearly,

$$P_t(A) = 1/2t + 1/2(1-t) = 1/2,$$

and therefore $P^*(A) = 1/2$. The conditional probabilities $P_t(A \mid B)$ and $P_t(A \mid \overline{B})$ are given by the expressions

$$P_t(A \mid B) = \frac{P_t(\{a\})}{P_t(\{a, c\})} = \frac{(1/2)t}{1/4 + 1/2t},$$

$$P_t(A \mid \overline{B}) = \frac{P_t(\{b\})}{P_t(\{b, d\})} = \frac{1/2(1-t)}{1/4 + 1/2(1-t)},$$

from which the lower bounds

$$P_*(A \mid B) = \inf_t P_t(A \mid B) = 0,$$

$$P_*(A \mid \overline{B}) = \inf_t P_t(A \mid \overline{B}) = 0,$$

are easily derived. It is clear, however, that

$$1/2 = P_*(A) > P_*(A \mid B) = P_*(A \mid \overline{B}) = 0$$

showing that the sandwich principle is violated even within the confines of conventional probability theory.

5.3. Other Ways to Spoil the Sandwich

Although such simple examples should suffice to dispose of concerns about spoiled sandwiches, we feel that Pearl's discussion of the problem deserves a more detailed analysis, mainly because of its philosophical implications for rational thinking. This is particularly important because loose use of such terms as "assured winnings," "support," or "belief" in the absence of a

sound, formal interpretive framework may quickly mislead those engaged in the comparison of alternative methodologies.

In an example called "the Peter, Paul, and Mary sandwich problem," Pearl presents a betting situation in which Mary prepares either a ham or a turkey sandwich, promising to pay Paul \$1000 should he guess correctly the type of sandwich that she has prepared. Not having a clue as to Mary's choice, Paul then flips a coin, guessing "ham" if the coin turns up heads and guessing "turkey" if it comes up tails. Paul, as Pearl notes, behaves like an "incurable Bayesian," reckoning that

$$\begin{aligned} P(\text{win}) &= P(\text{win} \mid \text{turkey})\, P(\text{turkey}) + P(\text{win} \mid \text{ham})\, P(\text{ham}) \\ &= P(\text{tails} \mid \text{turkey})\, \alpha + P(\text{heads} \mid \text{ham})(1 - \alpha) = 1/2 \end{aligned}$$

regardless of the value α of the probability that Mary has actually prepared a turkey sandwich. Thus, in spite of not being "assured" a win or having "supporting evidence," Paul can invoke the rationality (doubtful, as we already saw) of the sandwich principle and argue that he does not need to engage in unnecessary knowledge acquisition or experimentation [3]:

> If every possible outcome of an experiment would lead you to choose the same action, then you ought to choose that action without running the experiment.

From such an observation, Pearl proceeds to fault the philosophical underpinnings of the evidential reasoning approach, eventually going so far as to suggest that, should Bayesian orthodoxy be inapplicable, Dempster's formula —which, he freely admits, does not play any role in this example—be replaced by other formulas such as the well-known bounds recently rediscovered by Halpern and Fagin [12].

In the light of our previous example about the rather inconvenient ability of conventional probability families to spoil sandwiches, all of these pronouncements look increasingly suspicious. What, however, can we say is wrong? This question be answered in two equivalent ways.

We can say first, keeping ourselves at the informal discussion level, that, often, the experiments may interact with probabilities in complex ways that, obviously, Pearl has not considered. Nothing in Pearl's formalism suggests, for example, that the sandwich has already been prepared and that it may not be artfully substituted by Mary to ensure that Paul always loses, thus invalidating his hopes of having at least a 50% chance of winning.

The second, more formal, rendering of this observation is again based on the semantic model of Ruspini. In this, and in other similar problems, we have several agents that deliberate about the state of the world on the basis of their knowledge and their knowledge of the knowledge of others. If the unary operator **K** represents the state of knowledge of one of these agents, then, as

ON TRUTH, UTILITY, and SIMILARITY

Enrique H. Ruspini
Artificial Intelligence Center
SRI International
333 Ravenswood Avenue, Menlo Park, California 94025, U.S.A.

Abstract

We explore relations between the notions of utility, similarity, and multivalued truth that extend the logics of preference of Rescher. Emphasis is placed on the development of constructive procedures leading to solutions of control and decision-making problems that may be explicitly justified and explained.

Our departure point is the well-known interpretation, originally advanced by Bellman and Zadeh, of fuzzy sets on a solution space (i.e., a universe of possible worlds) as soft-constraint measures that quantify the relative desirability of alternative solutions. We propose also new procedures for the relaxation of problem-solving goals that are based solely on a conventional order relation defined among problem-solving constraints.

We explore mechanisms, based on multivalued logic, for the combination of desirability functions into more complex measures of solution quality and discuss differential preference relations between alternative state of affairs and, on the bases of certain formal relationships between them and desirability measures, we propose mechanisms for their logical combination.

We introduce and explore structures for the representation of ignorance about the utility of alternative solutions. These structures are based on generalizations of modal-logic notions of possible and necessary truth. Finally, we discuss relations between these concepts and similarity-based semantic models of fuzzy logic.

1. Introduction

In this paper we elaborate upon previous efforts [9] to develop a common framework to understand the aims, characteristics, and scope of application of several approximate-reasoning methodologies, including classical probability-based schemes, the Dempster-Shafer calculus of evidence, and fuzzy logic. In particular, this paper further explores the semantic underpinnings of fuzzy logic [8]. Specifically, we link the similarity-based structures discussed in our previous work to measures of utility that quantify the absolute or relative desirability of potential solutions to decision-making problems.

Our investigations were also strongly motivated by the requirement to produce a decision-making methodology that, going beyond the implementation of rule-based schemes, is capable of explaining the selection processes leading to a particular decision. We were specifically interested in the description of the rationale that permits to establish that some solution is worthier than another. Beyond simply stating that the value of some measure of quality is higher for some alternative w than for another alternative w', we tried to develop mechanisms to explicitly describe the reasons that make one solution preferrable to another.

Our investigations are based on ideas, first advanced by Rescher with his logics of preference [7],

which are based on the assignment of truth-values, in a [0,1]-scale, to sentences describing preference relations among uncertain outcomes of available decisions. This interpretation of truth values in terms of utilitarian concepts is similar in spirit to that originally proposed by Bellman and Zadeh [2] to describe the role of certain fuzzy sets in control and decision-making problems.

The results of our research, also concerned with the development of rational bases for the definition of similarity functions, owe much to previous efforts to introduce decision-theoretic concepts within the framework of fuzzy decision problems [3] and to studies on the structure and characteristics of fuzzy relations and operators [10,11].

The contribution described in this work may be summarized as the integration of these ideas, arising from developments on the theory of fuzzy sets and relations, with possible-world models that facilitate the description and interpretation of possibilistic structures, with a view to the advancement of methodologies for the qualitative modeling and analysis of complex systems. In this regard, we were strongly interested in the identification of control and planning policies that follow from rational analyses of approximate models of reality rather than from poorly understood human introspective processes.

Moreover, our efforts have been largely motivated by the need to integrate operation-research methods and logic-based procedures so as to develop techniques that produce decisions that may be justified and explained, i.e., to the description of the deliberations that led to the selection of such choices. To this end, we have also sought to clarify the logical relations that exist between individual measures of solution adequacy and global expressions that combine such measures into a single utility function on the basis of information about their relative degrees of importance.

Our formalism, like Rescher's, is based on the assignment of truth values to propositions of the form "It is desirable that p comes about." These values are directly derived from measures of the utility assigned by the decision-maker to such an outcome, or, in other words, the quantitative measures of the degree by which p is a "good thing."

In our approach, however, each of the propositions that defines the acceptability of a possible world as a solution is associated with a utility measure that provides a numerical ranking of all potential solutions, from the perspective of that restrictive statement, *all other things being equal*. Our measures are, therefore, expressions of the degree of adequacy of solutions from a limited perspective rather than measures of global relative desirability, regardless of context. In Rescher's approach, one such global measure is defined as an average of context-specific desirability values, an assumption that leads to the derivation of needlessly-restrictive properties for that function showing it to be similar to a probability distribution. In addition, to be able to represent degrees of knowledge about the potential utility of certain decisions, we introduce epistemic modalities [5] based on a generalization of well-known modal-logic concepts.

2. Possible Worlds and Desirabilities

Our model is based on the notion of *possible-world*, which may be informally characterized as the detailed description of any conceivable solution of a reasoning problem. In a propositional-logic framework, a possible world is an assignment of conventional truth-values (i.e., *true* or *false*) to the sentences that describe the possible state, behavior, or characteristics of a real-world system.

Any reasoning problem may be described as the determination of the set of possible worlds that complies with certain prescribed constraints, usually called the *knowledge* or the *evidence*. More generally we may think of such constraints, i.e., a set of restrictive propositions, as criteria to determine the acceptability or worthiness of any possible world as an answer to the problem. In what follows we

On Truth, Utility, and Similarty

observed before, our agent is always in one of three possible epistemological states with respect to the validity of a proposition p: Either he knows that p is true (denoted $\mathbf{K}\,p$) or he knows that p is false (denoted $\mathbf{K}\neg\, p$), or he may be ignorant of such truth (i.e., $\neg\mathbf{K} \wedge \neg\mathbf{K}\neg\, p$, denoted $\mathbf{I}\,q$).

In standard accounts, assuming that *knowledge* of the truth of one proposition does not affect the likelihood of *truth* of other propositions,[9] we are simply concerned with a single form of conditional probability: measuring the likelihood of p being true when q is true. In more complex epistemological situations, we may need to be concerned with such quantities as $P(\mathbf{K}\,p\,|\,\mathbf{K}\,q)$, $P(\mathbf{K}\,p\,|\,q)$, $P(\mathbf{K}\,p\,|\,\mathbf{I}\,q)$, and the like. In other words, Bel$(\,p\,|\,q)$ measures the support that knowledge of the truth of q provides to the truth of p, rather than the support provided by the truth of q to the truth of p.

In the Peter, Paul, and Mary sandwich problem, Pearl implicitly assumes that

$$P(\mathbf{K}_{\mathrm{MARY}}\text{heads}) = 0$$

$$P(\mathbf{K}_{\mathrm{MARY}}\text{tails}) = 0$$

$$P(\text{turkey}\,|\,\mathbf{I}_{\mathrm{MARY}}\text{heads}) = \alpha$$

$$P(\text{ham}\,|\,\mathbf{I}_{\mathrm{MARY}}\text{heads}) = 1 - \alpha$$

concluding correctly, by application of the total probability law, over the exhaustive and exclusive set of possibilities

$$\{\mathbf{K}_{\mathrm{MARY}}\text{heads},\ \mathbf{K}_{\mathrm{MARY}}\text{tails},\ \mathbf{I}_{\mathrm{MARY}}\text{heads}\}$$

that Paul has at least a 50% chance of winning.

This correct use of the total probability law does not mean that, by contrast, one should assume that the full extent of the conditional information by belief functions is limited to the conditional support functions

$$Bel(\,p\,|\,q) = P(\,p\,|\mathbf{K}q), \qquad Bel(\,p\,|\neg q) = P(\,p\,|\mathbf{K}\neg q)$$

as Pearl evidently does. In short, not knowing p is not the same as knowing $\neg\, p$. The example of the Peter, Paul, and Mary sandwich shows that one needs to consider states of ignorance that, when properly accounted for, spoil even the best-conceived principles of rationality.

To fully appreciate the complexity of the problem, suppose that we change Pearl's implicit assumptions, bringing the previously absent Peter into the scene as a spy acting on behalf of Mary. In this new scenario, still consistent

[9] The relations between knowledge and truth are more evident if "knowing" is thought of as sensing or observing, and if independence is understood as a lack of relationship between the errors of the sensors.

with Pearl's explicit statement of the problem. Peter, spying on Paul's coin-flipping experiment, alerts Mary, who, being rather artful and deft of hand, substitutes the sandwich so as to make sure that Paul always loses. In this case,

$$P(\text{ham} \mid \mathbf{K}_{\text{MARY}}\text{tails}) = 1, \qquad P(\text{turkey} \mid \mathbf{K}_{\text{MARY}}\text{heads}) = 1$$

and, most important

$$P((\mathbf{K}_{\text{MARY}}\text{heads}) \cup (\mathbf{K}_{\text{MARY}}\text{heads})) = 1,$$

that is, Mary is never ignorant as to what Paul will bet.

The Peter, Paul, and Mary sandwich example does not, in our view, invalidate the applicability of the evidential approach but rather highlights the need to make necessary discriminations between propositional truth, knowledge of that truth, and the interplay between such conditions that are likely to be glossed over by cursory analyses based on conventional approaches.

5.4. The Disagreeing Experts

Another common misunderstanding regarding the role of Dempster's combination formula is that provoked by an example of Zadeh [21], which is often described as an indication of theoretical inadequacy. This example concerns two reliable experts who assess, in a rather conflicting fashion, the likelihood of three noncompatible events A, B, and C as shown in Table 2. Representation of each of the expert's assessments as a mass distribution followed by their combination with Dempster's rule yields $P(B) = 1$, indicating that the "true" event is B, an alternative considered to be rather unlikely by either of the assessors.

Although this example is often quoted as an example of the failure of Dempster's rule, it is clear that each of the rows in Table 2 defines a conventional probability distribution, thus suggesting that the problem is likely to lie elsewhere. While one may be tempted to defend any method of evidence combination by saying that the evidence, however peculiar, indicates that Observer 1 is ruling out alternative C while Observer 2 is excluding alternative A, thus leaving only B as the sole possible answer, it is clear, upon further examination, that the rows of Table 2 cannot possibly be evaluations of the same probability distribution. If that were the case, then at least one of the

Table 2. Experts Disagree on the State of the World

Observer	$P(A)$	$P(B)$	$P(C)$
1	0.99	0.01	0
2	0	0.1	0.99

Table 3. The Berries Probability Distribution

Color	Size	Taste/Edibility	Probability
Red	Small	Good/edible	99/199
Blue	Large	Bad/edible	99/199
Red	Large	Poisonous	1/99

experts must be wrong, since there can be only one correct probability distribution, contradicting the assumption that they are both reliable.

Clearly, if the example is to make any sense—under any type of probabilistic interpretation—each row must correspond to a different conditional probability where the conditions correspond to different observations available to each expert. A simple example, suggested by a recent example used by Kyburg [22] to address other probabilistic reasoning issues, will help to clarify matters.

In this example we are being asked to reason, on the basis of available evidence, about the taste and edibility of certain berries that may be either small or large, red or blue, have good or bad taste, and be safe or poisonous to eat. We will assume that the berries in question are distributed according to the distribution shown in Table 3. If now a berry is picked up and found by an expert to be large, he will correctly conclude from such evidence that

$$P(\text{good} \mid \text{large}) = 0, \qquad P(\text{poisonous} \mid \text{large}) = 0.01,$$
$$P(\text{bad taste} \mid \text{large}) = 0.99.$$

Another expert, noticing that the berry is red, will conclude, on the other hand, that

$$P(\text{good} \mid \text{red}) = 0.99, \qquad P(\text{poisonous} \mid \text{red}) = 0.01,$$
$$P(\text{bad taste} \mid \text{large}) = 0.$$

Clearly, the evidential implications of these two separate observations are identical to the situation summarized in Table 2. Examination of Table 3, however, reveals that

$$P(\text{poisonous} \mid \text{red}, \text{large}) = 1$$

a correct solution that must be rationally expected from any reasoning method that purports to be valid.

The solution to the puzzle of the disagreeing experts lies in recognizing that there is, in fact, no disparity of opinion among them. Each is providing quantitative measures of likelihood with respect to *different* reference classes. Dempster's formula should never be applied to pool partial information about the same probability distribution. Furthermore, as shown by a sensitivity

analysis of the results of its application to the berries example, its use in situations where there is considerable disparity between reference classes (as suggested by the large normalization factor) should be discouraged on the basis of practical rather than conceptual considerations.

6. ON COMPLEXITY AND GENERALITY

The potential complexity of the belief function approach to represent and manipulate interval constraints on a family of probability distributions has often been mentioned as a handicap of the evidential reasoning methodology. In spite of such misgivings, two major empirical observations have indicated that the approach is applicable to a wide variety of practical problems.

First, our experience shows that, notwithstanding criticisms based on unrealistic worst-case scenarios, the approach is computationally efficient. In particular, we have found that representation of belief functions in terms of mass functions results in a storage and manipulation scheme that is both economical and easy to understand. In addition, we have successfully implemented tools, such as summarization and coarsening operators, that can be used effectively to limit representational complexity.

Second, our current functional operators have been chosen to guarantee that the manipulation of evidential knowledge results also in knowledge that can be represented in the evidential framework (i.e., the operators are closed).

The lack of generality of the belief function approach to represent general lower upper probability constraints is well known (Kyburg [23]). Our reliance on the methodology is primarily the result of practical considerations: Although we would prefer to manipulate more general constraints on probability values, compelling computational efficiency arguments force us to limit the scope of the problems considered to those capable of being at least approximately solved by a belief function treatment.

Being, in general, partial toward interpretations of evidential structures that are fully compatible with probability theory, our current research is being directed toward the development of more general, yet efficient, representation and manipulation methods.

Our current concerns with the manipulation of conditional and dependent evidence (i.e., the evidential counterpart of conditional probabilities) show, for example, that, for some important problems, the results of evidential combination fall outside the scope of its representational capabilities. In our experience, these methodological limitations are more worrisome than any of the supposedly paradoxical results arising from its misuse or its claimed lack of a decision-making apparatus.

Preliminary results (Ruspini [17]) indicate, on the other hand, that the belief function approach can be used to approximate the results of these evidential combination operations and that extended representation mechanisms (Spies

[24]) may yet be developed to treat more general evidential problems. This research also shows the basic errors inherent in criticisms that regard the belief function approach as a fully developed methodology incapable of sustaining further enhancement and modification. Because it has been studied in depth for only 15 years, its technological status is that of a young discipline, being capable both of enhancement on its own and of combination with other approaches to produce more general tools for probabilistic reasoning. Far from proving that we have reached a technological plateau, our investigations indicate that much is yet to be gained from such a development and integration process.

ACKNOWLEDGMENT

We thank Aviv Bergman, Aaron Bobick, Piero Bonissone, Luis de Campos, Allan Chateauneuf, Mike Clarke, Didier Dubois, Tom Garvey, Joe Halpern, Andrew Hanson, Y. T. Hsia, Jean-Yves Jaffray, Henry Kyburg, María Teresa Lamata, Abe Mamdani, Serafín Moral, Henri Prade, Alessandro Saffiotti, Glen Shafer, Prakash Shenoy, Philippe Smets, Marcus Spies, Nic Wilson, Llorenç Valverde, Leonard Wesley, John Yen, and Lotfi Zadeh for useful exchanges and insights.

This work was supported in part by the Defense Advanced Research Project Agency under contract N00039-88-C-0248, in part by the Air Force Office of Scientific Research under contract F49620-89-K-0001, and in part by the U.S. Army Research Office under contract DAAL03-89-K-0156. The views, opinions, and/or conclusions contained in this paper are those of the authors and should not be interpreted as representative of the official positions, decisions, or policies, either express or implied, of the supporting agencies or the United States Government.

References

1. Shafer, G., *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.
2. Lowrance, J., Strat, T., and Garvey, T., A framework for evidential-reasoning systems, *Proc. National Conference on A I*, 896–903, 1986.
3. Pearl, J., Reasoning with belief functions: an analysis of compatibility, *Int. J. Approx. Reasoning* 4; 363–390, 1990.
4. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Stat.* **38**; 325–339, 1967.
5. Ruspini, E. H., The logical foundations of evidential reasoning, Tech. Note 408, AI Center, SRI International, Menlo Park, Calif., 1987.

6. Hughes, G. E., and Creswell, M. J., *An Introduction to Modal Logic*, Methuen, New York, 1972.

7. Moore, R., Reasoning about knowledge and action, Tech. Note 191, SRI International, Menlo Park, Calif., 1980.

8. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, Calif., 1988.

9. Fagin, R., and Halpern, J. Y., Uncertainty, belief and probability, Res. Rep. RP 619/(60901), IBM Almaden Res. Center, Almaden, Calif., 1989.

10. Pearl, J., On probability intervals, *Int. J. Approx. Reasoning* 2; 211–216, 1988.

11. Smets, P., Belief functions, in *Positions and Debates in Nonstandard Logics*, (P. Smets, M.R.B. Clarke, C. Froidevaux, and J. Campbell, Ed.), Academic Press, New York, in press.

12. Halpern, J., and Fagin, R., Updating beliefs versus combining beliefs, Unpublished report, 1989.

13. Strat, T. M., Decision analysis using belief functions, *Int. J. Approx. Reasoning* 4; 391–418, 1990.

14. Wilson, N., Justification, computational complexity, and generalization of the Dempster combination formula, Unpublished report, 1989.

15. Diaconis, P., and Zabel, S. L., Some alternatives to Bayes's rule, in *Information Pooling and Group Decision Making* (B. Grofman and O. Guillermo, eds.), JAI Press, Greenwich, Conn., 1986, pp. 25–38.

16. Bar-Hillel, Y., and Falk, R., Some teasers concerning conditional probabilities *Cognition* 11; 109–122, 1982.

17. Ruspini, E. H., Approximate deduction in single evidential bodies, *Proceedings 1986 AAAI Uncertainty Workshop*, Philadelphia, 1986.

18. de Campos, L. M., Lamata, M. T., and Moral, S., The concept of conditional fuzzy measure, *Int. J. Intell. Syst.* (forthcoming).

19. Aleilunas, R. A., A new normative theory of probabilistic logic, *Proceedings of the 7th Biennial Conference of the Canadian Society for Computational Studies of Intelligence* (CSCSI-88), Edmonton, Alberta, 67–74, 1988.

20. Wilson, N., Private communication, 1990.

21. Zadeh, L. A., On the validity of Dempster's rule of combination of evidence, Memorandum No. UCB/ERL M79/24, Electronics Res. Lab., Univ. California, Berkeley, Calif., 1979.

22. Kyburg, H. E., Jerry's berries, Unpublished report, 1985.

23. Kyburg, H. E., Interval valued probabilities, in *Positions and Debates in Nonstandard Logics* (P. Smets, M.R.B. Clarke, C. Frodevaus, and J. Campbell, Eds.), Academic, New York, in press.

24. Spies, M., Combination of evidence with conditional objects and its application to cognitive modeling, in *Conditional Logic in Expert Systems* I. Goodmen and H. Nguyen, Eds., North-Holland, Amsterdam, 1990.

shall use the symbol $\mathcal{E}$ to denote the set of such evidentiary statements, or, equivalently, their logical conjunction.

Unlike conventional formalisms where restrictive propositions can only attain the classical truth-values *true* or *false*, [0, 1]-multivalued logics permit the definition of measures that gauge the relative adequacy of any solution. These elastic constraints describe the extent by which that solution meets some specific requirement, with a value of 1 being used to indicate that an alternative attains or exceeds target values or conditions, and a value of 0 utilized to describe complete failure to meet the standard.

It is important to remark, however, that these desirability measures rank solutions of a problem solely from the limited perspective of a single specific constraint and that they should not be confused with structures that represent the relative importance of each constraint or problem-solving goal, called "utilities of satisfaction" by Haddawy and Hanks [4].

2.1. Desirability Measures

In this section, we state, mainly for the sake of completeness, well-known interpretations of fuzzy sets [2] as elastic problem constraints, i.e., measures of relative desirability or adequacy from a specific viewpoint.

Definition: A *desirability measure* is a function $\mathbf{D} : \mathcal{U} \mapsto [0,1]$, i.e., a fuzzy set in the universe $\mathcal{U}$ of possible worlds.

Desirability measures clearly extend the conventional notion of "hard" or "crisp" constraint. The value $\mathbf{D}(w)$ attained by such a measure $\mathbf{D}$ for a particular solution w may also be considered to be the multivalued degree of truth of the proposition "The solution w is acceptable from the viewpoint of $\mathbf{D}$."

If w is a satisfactory solution from the viewpoint of a generalized requirement, which is expressed by means of the desirability measure $\mathbf{D}$, and from the viewpoint of another requirement, expressed by means of another desirability measure $\mathbf{D}'$, then it is clear that w should also be acceptable from the joint viewpoint of those two requirements. Well-known arguments [10] show that the desirability of a conjunction of two desirabilities $\mathbf{D}$ and $\mathbf{D}'$ is expressed by the relation

$$(\mathbf{D} \wedge \mathbf{D}')(w) = \mathbf{D}(w) \circledast \mathbf{D}'(w), \quad w \text{ in } \mathcal{U},$$

where $\circledast$ is a *triangular norm*.

Similarly, the desirability of the disjunction of two constraints $\mathbf{D}$ and $\mathbf{D}'$ can be seen to be given by

$$(\mathbf{D} \vee \mathbf{D}')(w) = \mathbf{D}(w) \oplus \mathbf{D}'(w), \quad w \text{ in } \mathcal{U},$$

where $\oplus$ is a *triangular conorm*.

In what follows we will assume that all triangular norms and conorms to be considered are continuous functions of their arguments since it is reasonable to require that the desirability of either the conjunction or the disjunction, respectively, of two arguments does not vary abruptly when there is a slight change of the desirabilities being combined.

Desirability measures that rank possible solutions by the degree by which they do *not* meet some constraint expressed by a desirability $\mathbf{D}$ are given by expressions of the form $\sim \mathbf{D}$, where $\sim$ is a *negation function*, i.e., a function $\sim$ from [0, 1] into [0, 1] that reverses the order of its argument, i.e., $\sim b \leq \sim a$ if $a \leq b$, and that, in addition, satisfies the relations $\sim 0 = 1$ and $\sim 1 = 0$.

The pseudoinverse $\oslash$ of a T-norm $\circledast$ is useful to generalize the implication operator $\rightarrow$ of classical logic by means of the expression $(\mathbf{D} \rightarrow \mathbf{D}')(w) = \mathbf{D}'(w) \oslash \mathbf{D}(w)$. The importance of pseudoinverses is related to the tautology $(\mathbf{D} \circledast (\mathbf{D} \rightarrow \mathbf{D}')) \rightarrow \mathbf{D}'$, which generalizes the classical modus ponens [10].

2.2. Constraint Importance

In a large number of situations, once several desirability measures $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n$, have been specified as the elastic constraints that restrict the acceptability of possible worlds as solutions to a reasoning problem, the goal of a problem-solving procedure may be simply described as the identification of possible worlds that are consistent with the evidence $\mathscr{E}$ and that, in addition, score sufficiently high values for the combined joint desirability measure

$$\mathbf{D}_1 \circledast \mathbf{D}_2 \circledast \ldots \circledast \mathbf{D}_n .$$

In other instances, however, the description of what is acceptable or not might involve the consideration of more complex logical expressions, as is the case in the famous example where a party host is told that it would desirable to invite John to a party, it would desirable to invite Mary to the same party, but that it would be rather unadvisable to have both present at the event. In that case, the desirability of a solution might be measured by the combined measure[1]

$$(\mathbf{D}_{\text{Mary}} \circledast \sim \mathbf{D}_{\text{John}}) \oplus (\sim \mathbf{D}_{\text{Mary}} \circledast \mathbf{D}_{\text{John}})$$

Furthermore, as pointed by various authors [1], requirements typically have unequal importance, a fact that has led to the consideration of broader connectives to derive a global measure of adequacy from the individual, goal-specific, measures $\mathbf{D}_i, 1 \leq i \leq n$. To some extent, the ability to use various triangular norms and conorms, simplifies the task of specifying how individual measures of adequacy may be traded off when assessing the overall worthiness of a potential solution. Thus, the minimum norm may be said to require minimum standards for each of the quality criteria being combined while the product norm provides for a reduction in the acceptability of one criteria given an appropriate increase in the quality of another. Such expressions, however, are based on utilization of symmetric functions that fail to adequately represent their different degrees of importance.

On the other hand, expressions suggested by optimization considerations such as

$$\alpha \mathbf{D}_1 + \beta \mathbf{D}_2 ,$$

which are based on the use of connectives having properties that lie between those of conjunctive and disjunctive operators [1] do not yield decision-making algorithms that are well-suited to explanatory analysis

Seeking to provide foundations for the derivation of explainable decision-making procedures, we have considered a number of typical situations arising in the context of control and decision problems where there exist basic requirements to develop procedures that selectively relax certain goals whenever all constraining statements cannot be met, usually expressed by unacceptably low values of the conjunction of the measures $\mathbf{D}_i$.

Although various extensions and enhancements are possible, the following situation should be sufficient to illustrate the nature of our results while keeping the complexity of our presentation to an acceptable level

In a decision-making problem, suppose that we seek to determine solutions u that satisfy a number of adequacy criteria $\mathbf{D}_i$, $0 \leq i \leq n$. Furthermore, assume that, whenever $i > j$, $\mathbf{D}_i$ is more important than $\mathbf{D}_j$, in the sense that if solutions having sufficiently high values for both $\mathbf{D}_i$ and $\mathbf{D}_j$ cannot be found, we would be willing to settle for those that have high enough values for $\mathbf{D}_i$. Furthermore, assume

[1] We are unconcerned here about the difficult problems faced by linguists who must translate typical utterances of such requirements, stated in language that loosely uses logical connectives, into its intended logical meaning

that conjunctions will be represented, as discussed above, by means of a triangular norm $\circledast$ having a pseudoinverse $\oslash$, and let the measures $\mathbf{E}_i$, $0 \leq i \leq n$, be defined by the expression

$$\mathbf{E}_i = \mathbf{D}_0 \wedge \mathbf{D}_1 \wedge \ldots \mathbf{D}_i, \quad 0 \leq i \leq n.$$

Note that $\mathbf{E}_n \leq \mathbf{E}_{n-1} \leq \ldots \mathbf{E}_1 \leq \mathbf{E}_0$. In addition, let $\mathcal{E}$ denote the set of possible worlds that describe certain minimum adequacy requirements imposed by available knowledge or evidence.

Our goal is to describe the overall adequacy of potential solutions to our problem by means of a fuzzy set $\mathcal{A}$, called the *acceptability* set, defined as the largest fuzzy subset of our universe of discourse that satisfies the following axioms:

1. $\mathcal{A} \Rightarrow \mathcal{E}$, i.e., all acceptable solutions must be consistent with the available evidence.

2. $\mathbf{E}_n \wedge \mathcal{E} \Rightarrow \mathcal{A}$, i.e., worlds in the most desirable class, if at all existent, are acceptable.

3. For all i such that $1 \leq i \leq n$ it is true that

$$\forall u, \quad \mathbf{E}_i(u) \wedge \mathcal{E}(u) \rightarrow \left[\, \forall u',\ \mathbf{E}_{i-1}(u') \wedge \mathcal{A}(u') \rightarrow \mathbf{E}_i(u') \,\right],$$

   i.e., if there are worlds that are desirable to the level i, then every acceptable world that is desirable at a level lower than i it is also acceptable to the level i or, informally, lower-quality alternatives are not acceptable if better choices are available.

Under these conditions, it is possible to show that the acceptability set $\mathcal{A}$ is given by the expression

$$\mathcal{A}(w) = \min \left[\mathcal{E}(w),\ \inf_{i \geq 1} \left[(\, \mathbf{E}_i(w) \oslash (\alpha_i \circledast \mathbf{E}_{i-1}(w')\,)\,) \right] \right],$$

where

$$\alpha_i = \sup_w \left[\mathbf{E}_i(w) \circledast \mathcal{E}(w) \right].$$

The general form of the expression suggests that the values α_i may be regarded as degrees of relative importance along lines suggested by Yager [12]. The above derivation, however, does not require the explicit specification of such quantities while making their value a function of the evidence $\mathcal{E}$.

2.3. Preference Relations

In a variety of situations it is easier to specify the degree by which some solution w is preferred to another solution w', from the viewpoint of some specific requirement, than to assign an absolute measure of adequacy toeither alternative.

This notion of relative desirability is formalized by functions of the form $\rho(w|w')$ that map pairs of possible worlds to numbers between 0 and 1 so as to quantify the extent by which a possible world w is preferred to another w', from the viewpoint of a particular constraint. To furnish ρ with the appropriate semantics we shall require it to satisfy the following conditions:

1. No resources should be spent to be in w if we are already in w.

2. If we are willing to spend resources to be in w when we are in w', then we should not spend any resources to be in w' if we were in w

3. The amount that we would be willing to pay to be in w when we are in w'' should be bound by above by a function of the amount that we would be willing to spend to be in w if we were in w' and of the amount that we would be willing to pay to be in w' if we were in w''.

These conditions readily lead to the following

Definition: A function ρ mapping pairs of possible worlds into numbers between 0 and 1 is called a $\oplus$-*preference* relation if and only if

1. $\rho(w|w) = 0$ for all w in $\mathcal{U}$.
2. If $\rho(w|w') > 0$, then $\rho(w'|w) = 0$ for all w and w' in $\mathcal{U}$.
3. For any possible worlds w, w' and w'' it is

$$\rho(w|w'') \leq \rho(w|w') \oplus \rho(w'|w'').$$

It is also easy to see that if ρ has the semantics of a relation representing graded preference, then $\oplus$ should be a conorm.

2.4. Relations between Desirabilities and Preferences

The combination and aggregation of preference relations is considerably more complex than that of desirability measures as, for example, the negation $\sim \rho$ of a preference relation ρ is not itself a preference relation. In order to develop an aggregation methodology, it is necessary first to study the relations that exist between both types of utilitarian measures.

The derivation of a $\oplus$-preference relation ρ_D from a desirability measure $\mathbf{D}$ is easily achieved by means of the pseudoinverse $\ominus$ of $\oplus$:

$$\rho_D\,(w|w') = \mathbf{D}(w) \ominus \mathbf{D}(w').$$

The inverse process of derivation of a unique desirability measure from a preference relation is, in general, not possible. One of several representation theorems of Valverde [11] exploiting in this case the identity

$$\rho(w|w') = \sup_{w'' \text{ in } \mathcal{U}} \left\{\rho(w|w'') \ominus \rho(w'|w'')\right\},$$

assures, however, that there is always a family $\{\mathbf{D}_\alpha\}$ of desirability measures such that

$$\rho(w|w') = \sup_{\alpha} \left\{\mathbf{D}_\alpha(w) \ominus \mathbf{D}_\alpha(w')\right\}.$$

The above representation has a most natural interpretation as the set of constraints (i.e., desirability measures) that are involved in the generalized order defined by a preference relation, i.e., the criteria that make a solution better than another. As it is often the case with conventional constraints, some of these generalized constraints may never be "active," being, in effect, superseded by more specific restrictions. For this reason, the above decomposition is never unique [6]. We may, however, always define a unique "canonical decomposition," which is suggested by the proofs of Valverde's theorems. We will call the family of desirability measures $\{\mathbf{D}_w\}$ defined by

$$\mathbf{D}_w(w') = \rho(w'|w), \qquad \text{for every } w \text{ in } \mathcal{U},$$

the *Valverde representation* of ρ.

Note that, although this relation essentially defines a mapping from every possible world w into a desirability measure $\mathbf{D}_w$, the collection of generating functions that is so defined may have a cardinality that is considerably smaller than that of $\mathcal{U}$. The question of whether there exists a unique desirability $\mathbf{D}$ measure that generates ρ, i.e., $\rho(w|w') = \mathbf{D}(w) \ominus \mathbf{D}(w')$, is, in view of the above comments, a matter of rather important practical significance, which was studied and solved by Jacas [6].

3. Combination of Preference Functions

The ability to express any preference function (i.e., relative adequacy of solutions) in terms of a collection of desirability measures (i.e., criteria for adequacy) also suggests a natural algebraic structure for preference relations.

Definition: Let ρ and ρ' be two preference relations in the universe of discourse $\mathcal{U}$. Furthermore, let $\{ \mathbf{D}_u \}$ and $\{ \mathbf{D}'_u \}$ be the Valverde representations of ρ and ρ', respectively. Then the conjunction and disjunction of ρ and ρ' are the preference functions, denoted $\rho \circledast \rho'$ and $\rho \oplus \rho'$, associated with the generating families $\{ \mathbf{D}_u \circledast \mathbf{D}'_u \}$, and $\{ \mathbf{D}_w \oplus \mathbf{D}'_w \}$, respectively. Furthermore, the complement of ρ is the preference relation $\sim \rho$ associated with the generating family $\{ \sim \mathbf{D}_u \}$. Finally, the implication preference $\rho \text{—} \rho'$ is the preference relation generated by the family $\{ \mathbf{D}_w \text{—} \mathbf{D}'_w \}$ of desirability measures.

4. Possibility and Necessity

It is often difficult to assess the adequacy of certain solutions even from the limited perspective of a single problem-solving goal. While steering a mobile robot around an obstacle, for example, it is hard to determine if a particular move is preferrable to another from the viewpoint of a maneuver to be performed later at a remote location.

Modal logics [5], by introduction of notions of possible and necessary truth, permit to represent states of ignorance about the potential truth of the different statements that are being reasoned about. In the formalism presented in this paper, where restrictive propositions have been generalized as relative measures of solution adequacy, the role of the necessity and possibility operators of modal logic is replaced by lower and upper bounds for measures of desirability and preference.

We will say, therefore, that a function $\mathbf{N}_D$ mapping possible worlds w into values between 0 and 1 is a ***necessary desirability distribution*** for a desirability measure $\mathbf{D}$ if $\mathbf{N}_D(w) \leq \mathbf{D}(w)$ for all w in $\mathcal{U}$. Similarly, we will say that $\mathbf{\Pi}_D$ is a ***possible desirability distribution*** for $\mathbf{D}$ if $\mathbf{D}(w) \leq \mathbf{\Pi}_D(w)$ for all w in $\mathcal{U}$.

The following results permit to manipulate necessary and possible desirabilities along lines that generalize similar derivation procedures for conventional modal logic.

(a) If $\mathbf{N}_{\sim D}$ is a necessary desirability for the complement $\sim \mathbf{D}$ of $\mathbf{D}$, then $\sim \mathbf{N}_{\sim D}$ is a possible desirability for $\mathbf{D}$. Similarly, if $\mathbf{\Pi}_{\sim D}$ is a possible desirability for the complement $\sim \mathbf{D}$ of $\mathbf{D}$, then $\sim \mathbf{\Pi}_{\sim D}$ is a necessary desirability for $\mathbf{D}$. These relations are the generalization of the well-known duality relations $\neg\mathbf{N}\neg p \equiv \mathbf{\Pi} p$ and $\neg\mathbf{\Pi}\neg p \equiv \mathbf{N} p$.

(b) If $\mathbf{N}_D$ and $\mathbf{N}_{D'}$ are necessary desirabilities for $\mathbf{D}$ and $\mathbf{D}'$, respectively, then $\mathbf{N}_D \circledast \mathbf{N}_{D'}$ and $\mathbf{N}_D \oplus \mathbf{N}_{D'}$ are necessary desirabilities for $\mathbf{D} \circledast \mathbf{D}'$ and $\mathbf{D} \oplus \mathbf{D}'$, respectively. A similar statement holds for possible desirabilities.

(c) If $\mathbf{N}_D$ is a necessary desirability for $\mathbf{D}$ and if $\mathbf{\Pi}_{D'}$ is a possible desirability for $\mathbf{D}'$, then $\mathbf{N}_D \div \mathbf{\Pi}_{D'}$ is a necessary desirability for $\mathbf{D}' \text{—} \mathbf{D}$. A dual statement also holds for possible desirabilities.

Bounds, called *necessary* and *possible preference functions*, may also be introduced to represent ignorance about relative preference between solutions. Rules for their manipulation, however, are considerably more complex than those for their desirability counterparts. A rather straightforward consequence, nonetheless, of the definition of preference functions is that if if $\mathbf{N}_D$ and $\mathbf{\Pi}_D$ are necessary and possibility desirability distributions for a desirability measure $\mathbf{D}$, then the functions defined by the expressions

$$\mathbf{N}_\rho(w|w') = \mathbf{N}_D(w) \ominus \mathbf{\Pi}_D(w'), \quad \text{and} \quad \mathbf{\Pi}_\rho(w|w') = \mathbf{\Pi}_D(w) \ominus \mathbf{N}_D(w'),$$

are necessary and possible preferences for $\rho_D(w|w') = \mathbf{D}(w) \ominus \mathbf{D}(w')$.

It should be also clear that necessary and possible preference functions can always be chosen to satisfy the first two properties (generalized nonreflexivity and antisymmetry) of the definition of preference function. Less obvious is the fact that a possible preference function may always be selected to satisfy the third (or transitive) property. Since then such possible preference relation will be itself a preference relation, it may be represented by a family $\hat{\mathbf{D}}_w$ of desirability measures that is related to the Valverde representation $\mathbf{D}_w$ of ρ by the inequality $\mathbf{D}_w \leq \hat{\mathbf{D}}_w$.

5. Preference, Similarity, and Fuzzy Logic

A recent semantic model of the author [8] presented a rationale for the interpretation of the possibilistic structures of fuzzy logic and for its major rule of derivation on the basis of similarity relations between possible worlds. Similarity relations S assign a value $S(w, w')$ between 0 and 1 to every pair of possible worlds w and w' in such a way that

1. $S(w, w) = 1$ for all possible worlds w,
2. $S(w, w') = S(w', w)$ for all possible worlds w and w', and
3. $S(w, w') \leq S(w, w'') \circledast S(w'', w')$ for all possible worlds w, w' and w'', where $\circledast$ is a T-norm

Two possible worlds w and w' may be considered similar if, from the perspective of all constraints defining a problem, the solutions that they represent have close desirability values. This statement, reflected by the well known relation

$$S(w, w') = \min\left(\sim \rho(w|w'), \sim \rho(w'|w)\right),$$

permits derivation of a similarity relation from a preference relation. Extensions of the notion of similarity to allow definition of bounds for the resemblance between subsets of possible worlds, called *degree of implication* and *degree of consistence*, play an essential role in the interpretation of the possibility distributions of fuzzy logic.

Acknowledgements

This work was supported in part by the United States Army Research Office under Contract No. DAAL03-89-K-0156, by the United States Air Force Office of Scientific Research under Contract No. F49620-91-C-0060, and in part by a contract with the Laboratory for International Fuzzy Engineering Research. The views, opinions and/or conclusions contained in this note are those of the author and should not be interpreted as representative of the official positions, decisions, or policies, either express or implied, of his sponsors.

The author wants to thank F. Esteva, L. Godo, N. Helft, D. Israel, J. Jacas, J. Lowrance, R. Perrault, E. Trillas, L. Valverde, and L. Zadeh for many useful insights and exchanges.

This paper is an extension of previous work presented at the Conference on Uncertainty in Artificial Intelligence, Los Angeles, California, 1991.

References

[1] N. Piera and J. Aguilar Martin. *Mixed Connectives of Lineal Compensation*. Report LAAS 88031, Laboratoire d'Automatique et d'Analyse des Systemes, Toulouse, France, 1988.

[2] R.E. Bellman and L.A. Zadeh. Decision-making in a fuzzy environment. *Management Science*, 17:B141–B164, 1980.

[3] D Dubois and H. Prade. Criteria aggregation and ranking of alternatives in the framework of fuzzy set theory. In H.J. Zimmerman, L.A. Zadeh, and B.R. Gaines, editors, *Fuzzy Sets and Decision Analysis*, pp. 209–240, North Holland, Amsterdam, 1984.

[4] P. Hadawy and S. Hanks. Issues in Decision-Theoretic Planning: Symbolic Goals and Numeric Utilities. In K.P. Sycara, editor, *Proceedings of a Workshop on Innovative Approaches to Planning, Scheduling, and Control*, pp. 48–58. San Diego, California, 1990.

[5] Hughes, G.E. and M.E. Creswell, *An Introduction to Modal Logic*. Methuen, London, 1968.

[6] J. Jacas Moral. Contribució a l'estudi de les relacions d' indistingibilitat i a les seves aplicacions als processos de clasificació. Ph. D. Thesis. Polytechnic University of Barcelona, 1987.

[7] N. Rescher. Semantic Foundations for the Logic of Preference. In N. Rescher, editor, *The Logic of Decision and Action*, Pittsburgh, 1967.

[8] E.H. Ruspini. On the Semantics of Fuzzy Logic. *Int.J. Approximate Reasoning*, 5, 45–88, 1991.

[9] E.H. Ruspini. Approximate Reasoning: Past, Present, Future. *Information Sciences*, forthcoming, 1991. Also published as Technical Note No. 492, Artificial Intelligence Center, SRI International, Menlo Park, California, 1990.

[10] E. Trillas and L. Valverde. On mode and implication in approximate reasoning. In M.M. Gupta, A. Kandel, W. Bandler, J.B. Kiszka, editors, *Approximate Reasoning and Expert Systems*, North Holland, Amsterdam, 157–166, 1985.

[11] L. Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17: 313-328, 1985.

[12] R.W. Yager. A new methodology for ordinal multiple decisions based on fuzzy set theoty. *Decision Science*, 12:589-600, 1981

Blending Reactivity and Goal-directness in a Fuzzy Controller

Procs. of the Second IEEE Conf. on Fuzzy Systems
San Francisco, CA, March 1993, pp. 134-139

Blending Reactivity and Goal-Directedness in a Fuzzy Controller

Alessandro Saffiotti* **Enrique H. Ruspini** **Kurt Konolige**
Artificial Intelligence Center, SRI International
Menlo Park, CA 94025, U.S.A.
saffiotti@ai.sri.com

***Abstract*— Controlling the movement of an autonomous mobile robot requires the ability to pursue strategic goals in a highly reactive way. We describe a fuzzy controller for such a mobile robot that can take abstract goals into consideration. Through the use of fuzzy logic, reactive behavior (e.g., avoiding obstacles on the way) and goal-oriented behavior (e.g., trying to reach a given location) are smoothly blended into one sequence of control actions. The fuzzy controller has been implemented on the SRI robot Flakey.**

I. Introduction

Autonomous operation of a mobile robot in a real environment poses a series of problems. In the general case, knowledge of the environment is partial and approximate; sensing is noisy; the dynamics of the environment can only be partially predicted; and robot's hardware execution is not completely reliable. Though, the robot must take decisions and execute actions at the time-scale of the environment. Classical planning approaches have been criticized for not being able to adequately cope with this situation, and a number of reactive approaches to robot control have been proposed (e.g., [Firby, 1987; Kaelbling, 1987; Gat, 1991]), including the use of fuzzy control techniques (e.g., [Sugeno and Nishida, 1985; Yen and Pfluger, 1992]). Reactivity provides immediate response to unpredicted environmental situations by giving up the idea of reasoning about future consequences of actions. Reasoning about future consequences (sometimes called "strategic planning"), however, is still needed in order to intelligently solve complex tasks (e.g., by deciding not to carry an oil lantern downstairs to look for a gas leak [Firby, 1987].)

One solution to the dual need for strategic planning and reactivity is to adopt a two-level model: at the upper level, a planner decides a sequence of abstract goals to be achieved, based on the available knowledge; at the lower level, a reactive controller achieves these goals while dealing with the environmental contingencies. This solution requires that the reactive controller be able to simultaneously satisfy strategic goals coming from the planner (e.g., going to the end of the corridor), and low-level "innate" goals (e.g., avoiding obstacles on the way). A major problem in the design of such a controller is how to resolve conflicts between simultaneous goals.

In this paper, we describe a reactive controller for an autonomous mobile robot that uses fuzzy logic for trading off conflicting goals. This controller has been implemented on the SRI robot Flakey, and its performance demonstrated at the first AAAI robot competition, where Flakey finished second [Congdon *et al.*, 1993]. The formal bases for the proposed controller have been set forth by Ruspini [Ruspini, 1990; Ruspini and Ruspini, 1991; Ruspini, 1991a] after the seminal works by Zadeh (e.g., [Zadeh, 1978]). In a nutshell, each goal is associated with a function that maps each perceived situation to a measure of desirability of possible actions from the point of view of that goal. The notion of a *control structure* is used for introducing high-level goals into the fuzzy controller. Intuitively, a control structure is an object in the robot's workspace, together with a desirability relation: typical control structures are locations to reach, walls to follow, doors to enter, and so on. Each desirability function induces a particular behavior — one obtained by executing the actions with higher desirability. Behaviors induced by many simultaneous goals can be smoothly blended by using the mechanisms of fuzzy logic. In particular, reactive and goal-oriented behaviors are blended in this way into one sequence of control actions.

The next section gives a brief overview of Flakey. Section III sketches the architecture of the controller, and describes the way behaviors are implemented, and how they are blended together. Section IV deals with the introduction of high-level goals into the reactive controller. Section V discusses the results, and concludes.

*On leave from Iridia, Université Libre de Bruxelles, Brussels, Belgium.

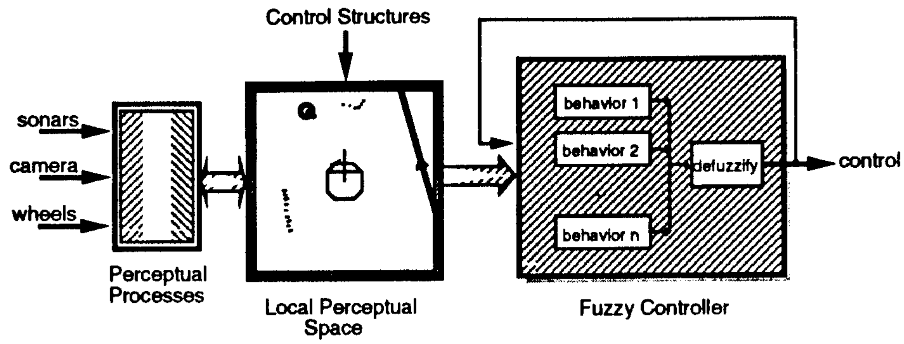


Figure 1: Architecture of Flakey (partial)

II. THE MOBILE ROBOT TEST-BED

Flakey is a custom-built mobile robot platform approximately 1 meter high and .6 meter in diameter for use in an indoor environment. There are two independently-driven wheels, one on each side, giving a maximum linear velocity of about .5 meters/sec. Flakey sensors include a ring of 12 sonars, giving information about distances of objects up to about 2 meters; wheel encoders, providing information about current linear and rotational velocity; and a video camera, currently used in combination with a laser to provide dense depth information over a small area in front of Flakey. On-board computers are dedicated to low-level sensor interpretation, motor control, and radio communication with an off-board Sparc station. Though it is possible to run the high level interpretation and control processes on board, they are normally run remotely for programming convenience.

Figure 1 illustrates the part of Flakey's architecture that is relevant to the controller. The sensorial input is processed by a number of interpretation processes at different levels of abstraction and complexity, and the results of interpretation are stored in the *local perceptual space* (LPS). The LPS represents a Cartesian plane centered on Flakey where all the information about the vicinity of Flakey is registered. In Figure 1, points corresponding to surfaces identified by the sonars and the camera are visible in the LPS — Flakey is the the octagon in the middle of the LPS, in top-view. The other objects in the LPS are "artifacts" associated to *control structures*, and are discussed in Section V. The content of the LPS constitutes the input to the controller: this checks its input and generates a control action every 100 milliseconds.

III. REACTIVE FUZZY CONTROLLER

The fuzzy controller is centered on the notion of *behavior*. Intuitively, a behavior is one particular control regime that focuses on achieving one specific, predetermined goal (e.g., avoiding obstacles). Hence, we can think of a behavior as a mapping from configurations in the LPS to actions to perform. More precisely, and following [Ruspini, 1991b], we say that each behavior B is associated with a desirability function

$$Des_B : \text{LPS} \times \text{Control} \rightarrow [0,1]$$

that measures, for each configuration s of the LPS and value c of a control variable, the desirability $Des_B(s,c)$ of applying control values c in the situation s *from the point of view of B*. Equivalently, we can say that Des_B associates each situation s with the fuzzy set $\widetilde{C}$ of control values characterized by the membership function $\mu_{\widetilde{C}}(c) = Des_B(s,c)$. Notice that in general, c is a n-dimensional vector of values for all the control variables; in the case of Flakey, the control variables include linear acceleration and turning angle.

In practice, each behavior is implemented by a fuzzy machine structured as shown in Figure 2. The *fuzzy state* is a vector of fuzzy variables (each having a value in $[0,1]$) representing the truth values of a set of fuzzy propositions of interest (e.g., "obstacle-close-on-left"). At every

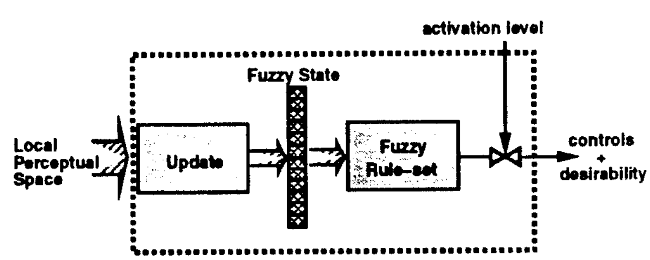


Figure 2: Implementation of a behavior.

cycle, the **Update** module look at the (partially) interpreted perceptual input stored in LPS, and produces a new fuzzy state. The **Fuzzy Rule-Set** module contains a set of fuzzy rules of the form "If A then c" where A is a fuzzy expression composed by predicates in the fuzzy states plus the fuzzy connectives AND, OR and NOT; and c is a vector of values for the control variables. Max, min, and complement to 1 are used to compute the truth value of disjunction, conjunction and negation, respectively. An example of a control rule is:

```
IF obstacle-close-in-front
AND NOT obstacle-close-on-left
THEN turn -6 degrees
```

Each "If A then c" rule computes the degree of desirability of applying control value c as a function of the degree at which the current state happens to be similar to A. The outputs of all the rules in a rule-set are unioned using the max T-conorm: the function computed in this way is meant to provide an approximation of the Des_B function above.[1] This desirability function is fed to the **Defuzzify** module for computing one single control value. We presently do defuzzification according to the centroid approach: the resulting control value is given by

$$\frac{\int c\, Des_B(c)\, dc}{\int Des_B(c)\, dc}.$$

As shown in Figure 1 above, many behaviors can be simultaneously active in the controller, each aiming at one particular goal — e.g., one for avoiding obstacles; one for keeping a constant speed; one for heading toward a beacon; etc. Correspondingly, many instances of the fuzzy machine depicted in Figure 2 simultaneously run in the controller, each one implementing one behavior's desirability function. All these desirability functions are merged into a composite one by the max T-norm; the defuzzification module converts the resulting tradeoff desirabilities into one crisp control decision. Care must be taken, however, of possible conflicts among behaviors aiming at different, incompatible goals. These conflicts would result in desirability functions that assign high values to opposite actions: simple T-norm composition should not be applied in these cases. The key observation here is that each behavior has in general its own *context* of applicability. Correspondingly, we would like that the impact of the control actions suggested by each behavior be weighted according to that behavior's degree of applicability to the current situation. For instance, the actions proposed by the obstacle avoidance behavior should receive higher priority when there is a danger of collision, at the expense of the other, concurrent behaviors. In order to do this, the output of each rule-set is discarded by the value of the corresponding *activation level*: typically, the activation level is represented by some variable in the fuzzy state. This corresponds to arbitrate the relative dominance of different behaviors by a set of meta-rules of the form

$$\text{IF } A' \text{ THEN activate\_behavior } B \tag{1}$$

where A' is a LPS configuration. Notice that this solution is formally equivalent to transforming each rule "If A then c" in B into a rule "If A' and A then c" (see [Berenji *et al.*, 1990] for a similar approach to conflict resolution.)

As an example, consider the way Flakey "wanders" around. In the wandering mode, three behaviors coexist in the controller: AVOID-OBSTACLES, AVOID-COLLISIONS and GO-FORWARD. GO-FORWARD just keeps Flakey going at a fixed velocity, given as a parameter. AVOID-OBSTACLES looks at the last 5 seconds' sonar readings in the LPS, and guides Flakey away form occupied areas. AVOID-COLLISIONS looks at the nearest sonar readings and proposes drastic actions (immediate stop and turn) when a serious risk of collision is detected. The activation levels of AVOID-OBSTACLES and AVOID-COLLISIONS are given by the fuzzy state variable "approaching-obstacle"; the complement of this value gives the activation level for GO-FORWARD. The visual result for an external observer is that Flakey "follows its nose", while smoothly turning away from obstacles as it approaches them.

IV. BEYOND PURE REACTIVITY

The behaviors discussed in the previous section are purely reactive: at each cycle, Flakey selects an action solely on the basis of the current state of the world as perceived by its sensors and represented in the local perceptual space. Engaging into more purposeful activities than just wandering around requires more than pure reactivity: we need to take explicit goals into consideration. For example, we may want Flakey to reach a given position at a given velocity, and still (reactively) avoid the obstacles on the way.[2]

In our approach, a goal is represented by a *control structure*. Intuitively, a control structure is virtual object (an *artifact*) that we put in the LPS, associated with a behavior that encodes the way to react to the presence of this object. For example, a "control-point" is a marker for a (x, y) location, together with a heading and a velocity: the associated behavior GO-TO-CP reacts to the presence of a control point in the LPS by generating the commands to reach that position, heading and velocity. In Figure 1 there are two artifacts: a control point to reach (left), and a wall to follow (right).

[1] See [Ruspini, 1991a; Ruspini, 1991b] for an account of fuzzy logic and fuzzy control in terms of similarity and desirability measures, and the use of T-norms and T-conorms in this context.

[2] Reactive behaviors are also associated with (innate) goals, hardwired in the definition of the behavior. We are now interested in dynamically assigning specific strategic goals to Flakey.

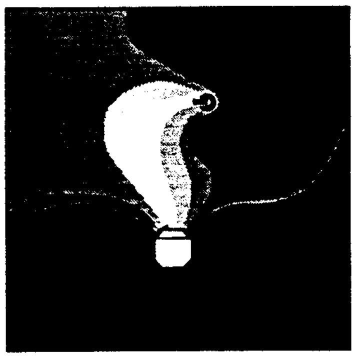


Figure 3: Path families generated by actions with increasing values of Des_{FS}.

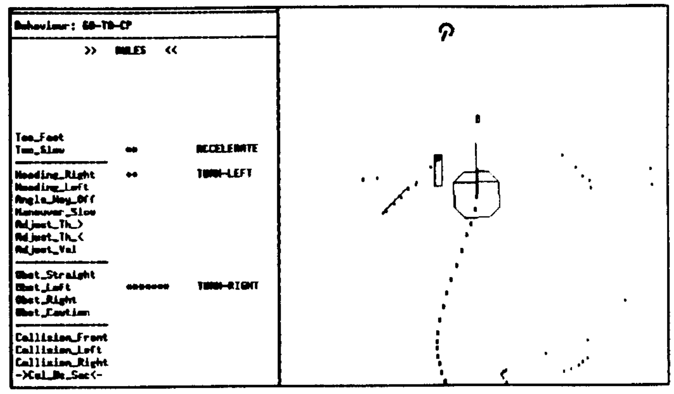


Figure 4: A snapshot of Flakey's control window while achieving a control point.

More precisely, a control structure is a pair

$$S = \langle Q_S, R_S \rangle,$$

where Q_S is an artifact, and R_S is a fuzzy relation between the position of Flakey and that of the artifact.[3] Such a control structure implicitly defines a goal: the goal to achieve, and maintain, the given relation between Flakey and the artifact Q_S. Intuitively, if Q_S is at position q, $R_S(q,p)$ says how much a position p of Flakey satisfies this goal. If the position of Flakey is such that $R_S(q,p) = \alpha$, we say that the control structure S is satisfied to the degree α.

The R_S relation induces a desirability function Des_S in the following way. Given the set P of possible positions of Flakey, and the set C of possible control values, let $\mathrm{Exec}(p,c)$ denote the new position reached by applying control c from position p. Then, the desirability *from the viewpoint of the control structure* S of executing c when Flakey is at position p (and Q_S is at q) is given by

$$Des_S(q,p,c) = R_S(q, \mathrm{Exec}(p,c))$$

However, not all positions are equally reachable by Flakey: moving to certain positions will require more effort (changes in velocity and/or direction, time, etc.) than moving to others. To account for this, we consider a second desirability function Des_F: $Des_F(p,c)$ measures the desirability *from the viewpoint of Flakey's motion capabilities* of executing control action c when Flakey is at position p. The desirability of control actions from the joint viewpoints of feasibility for Flakey, and effectiveness with respect to the control structure S, is measured by the combination $Des_S \otimes Des_F$ (where $\otimes$ is a T-norm). Figure 3 illustrates one such combined desirability function: here, S is a control point, represented by the semi-circle near the top (the "tail" indicates the desired entry orientation.) The fading from black to white illustrates the increase in the value of $Des_S \otimes Des_F$ for some families of possible paths.

We have already seen how Des_S can induce, for each LPS configuration, a fuzzy set of possible controls. As we did in the case of reactive behaviors, we approximate this fuzzy set using rules on the form "If A then c". The only difference is that A now refers to artifacts rather than to sensorial input.[4] We have designed sets of rules for many "purposeful" behaviors, including going to a x, y position; achieving a control point; following a wall; crossing a door; and so forth. Each ruleset consists of a small number (four to eight) of rules. Purposeful behaviors can coexist with other behaviors, either purposeful and reactive: the context-dependent blending of behaviors explained above provides arbitration and guarantees the smooth integration of directed activities and reactivity.

Figure 4 exemplifies the performance of the integration. The picture shows Flakey's control window during an actual run: on the right is Flakey's local perceptual space. Flakey sits in the middle of the window, pointing upwards; the small points all around mark sonar readings, indicating the possible presence of some object; the rectangle on the left of Flakey highlight a dangerously close object. The window on the left lists all the currently active rules, grouped into rule-sets: topmost, the rules for the GO-FORWARD behavior; below, those for GO-TO-CP, for AVOID-OBSTACLES, and for AVOID-COLLISIONS. In the shown situation, Flakey is going too slow and heading right of the CP: hence, some desirability is given to the

[3] Positions are actually points in a (x, y, θ, v) 4-D space.

[4] Alternatively, these rules can be thought as responding to input from a "virtual sensor" that senses the position of an artifact.

accelerate and the *turn-left* actions. However, the close obstacle on the left causes the activation level of the AVOID-OBSTACLES behavior to be high, at the expenses of the other behaviors; hence, the *turn-right* action suggested by AVOID-OBSTACLES receives high total desirability (as indicated by the 7 stars). The small box in front of Flakey indicates the resulting turning control — some degrees on the right. The overall result of the blending is that Flakey makes its way among obstacles while *en route* to achieving the position and bearing of the given control point. The smoothness of the movement in evident in the wake of small boxes that Flakey left behind it (one box per second). Flakey's speed was between 200 and 300 mm/sec.

One word is worth spent on the problem of local minima, ubiquitous in approaches to robot navigation based on local combination of behaviors [Latombe, 1991]. The problem is illustrated in Figure 5 (top): the robot needs to mediate the tendency to move toward the goal, and the tendency to stay away from the obstacle. A straightforward combination of these two opposite tendencies (whether they are described by desirability measures, potential fields [Khatib, 1986], motor schemas [Arkin, 1990], or other) may result in the production of a zone of local equilibrium (local minimum): when coming from the left edge, the robot would be first attracted and then trapped into this zone. By using meta-rules like the 1 above to reason about the relative importance of goals, our context-dependent blending of behaviors provides a way around this problem. Figure 5 shows the path followed by Flakey in a simulated run (top), and the corresponding activation levels of the KEEP-OFF and REACH behaviors (bottom). In (a), Flakey has perceived the obstacle; as the obstacle becomes nearer, the KEEP-OFF behavior becomes more active, at the expenses of the REACH behavior. In this way, the "attractive power" of the goal is gradually shaded away by the obstacle, and Flakey responds more and more to the obstacle-avoidance suggestions alone. The REACH behavior re-gains importance, however, as soon as Flakey is out of danger (b).

V. CONCLUSIONS

We have defined a mechanism based on fuzzy logic for blending multiple behaviors aimed at achieving different, possibly conflicting goals. Goals are either built-in, as in most fuzzy controllers, or dynamically set from outside the controller. Typically, the built-in goals correspond to reactive behaviors (like avoiding collisions), while the dynamic ones are strategic goals communicated by a planner. Context-dependent blending of behaviors ensures that strategic goals be achieved as much as possible, while maintaining a high reactivity.

Our behavior blending mechanism has been originally inspired to the technique proposed by Berenji et al.

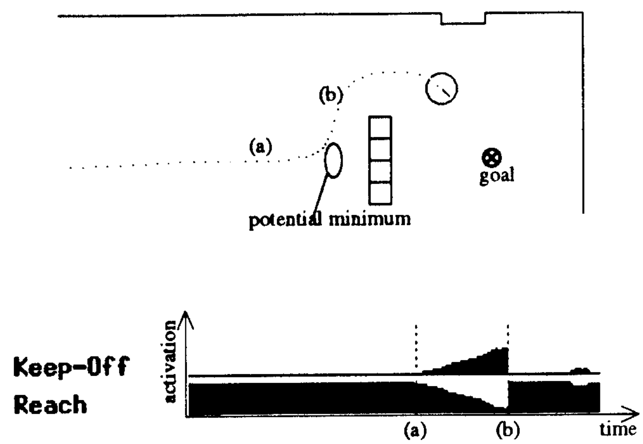


Figure 5: How context-dependent blending of behaviors avoids potential local minima.

[Berenji *et al.*, 1990] for dealing with multiple goals in fuzzy control. There are however two important differences: first, our context mechanism dynamically modifies the degrees of importance of each goal; second, we allow the introduction of high-level, situation-specific goals in the controller.

From another perspective, the work presented here fits in the tradition of the "two level" approaches to robot control, where a strategic planner is used to generate guidelines to a reactive controller (e.g., [Arkin, 1990; Payton *et al.*, 1990; Gat, 1991]). In our case, a plan consists in a sequence of control structures. For example, a plan to exit building E could consist in three successive *corridors* to follow, one *control point* in the entrance hall close to the door, and the exit *door* itself. The context of applicability of each control structure is used to decide when each control structure becomes relevant. (see [Saffiotti *et al.*, 1993; Saffiotti, 1993] for more on this issue). We believe that having based our architecture on fuzzy logic results in improved robustness (e.g., more tolerance to sensor noise and knowledge imprecision), while granting a better understanding of the underlying mechanisms.

Finally, many current approaches to robot control deal with multiple goals using the so-called "potential fields" method [Khatib, 1986]: goals are represented by pseudo-forces, which may be thought of as representatives of most desirable behavior from that goal's viewpoint. These optimal forces are then combined, as physical vectors, to produce a resultant force that summarizes their joint effect. In our approach, by contrast, the goals' desirability functions, rather than a summary description, are combined into a joint desirability function, from which a most desired tradeoff control is extracted. Moreover, this com-

bination takes behaviors' context of applicability into account; this provides a key to eliminate the local minima arising from the combination of conflicting goals.

The technique proposed in this paper has been implemented in the SRI mobile robot Flakey, resulting in extremely smooth and reliable movement. The performance of Flakey's controller has been demonstrated at the first AAAI robot competition in San Jose, CA [Congdon *et al.*, 1993]. Flakey accomplished all the given tasks while smoothly getting around obstacles (whose positions were not known beforehand) and people, and placed second behind Michigan University's CARMEL. Flakey's reliable reactivity is best summarized in one judge's comment: "Only robot I felt I could sit or lie down in front of." (What he actually did!)

Acknowledgments John Lowrance, Daniela Musto, Karen Myers and Leonard Wesley contributed to the development of the ideas presented in this paper. Nicolas Helft implemented a first version of Flakey's controller.

The first author has been supported by a grant from the National Council of Research of Italy. Research performed by the second author leading to the conceptual structures used in the autonomous mobile vehicle controller was supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620-91-C-0060. Support for Kurt Konolige came partially from ONR Contract No. N00014-89-C-0095. Additional support was provided by SRI International.

REFERENCES

[Arkin, 1990] Arkin, Ronald C. 1990. The impact of cybernetics on the design of a mobile robot system: a case study. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6):1245-1257.

[Berenji *et al.*, 1990] Berenji, H.; Chen, Y-Y.; Lee, C-C.; Jang, J-S.; and Murugesan, S. 1990. A hierarchical approach to designing approximate reasoning-based controllers for dynamic physical systems. In *Procs. of the 6th Conf. on Uncertainty in Artificial Intelligence*, Cambridge, MA.

[Congdon *et al.*, 1993] Congdon, C.; Huber, M.; Kortenkamp, D.; Konolige, K.; Myers, K.; and Saffiotti, A. 1993. CARMEL vs. Flakey: A comparison of two winners. *AI Magazine*. To appear.

[Firby, 1987] Firby, J. R. 1987. An investigation into reactive planning in complex domains. In *Procs. of the AAAI Conf.*

[Gat, 1991] Gat, E. 1991. *Reliable Goal-Directed Reactive Control for Real-World Autonomous Mobile Robots.* Ph.D. Dissertation, Virginia Polytechnic Institute and State University.

[Kaelbling, 1987] Kaelbling, L. P. 1987. An architecture for intelligent reactive systems. In Georgeff, M.P. and Lansky, A.L., editors 1987, *Reasoning about Actions and Plans*. Morgan Kaufmann.

[Khatib, 1986] Khatib, O. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research* 5(1):90-98.

[Latombe, 1991] Latombe, J.C. 1991. *Robot Motion Planning.* Kluver Academic Publishers, Boston, MA.

[Payton *et al.*, 1990] Payton, D. W.; Rosenblatt, J. K.; and Keirsey, D. M. 1990. Plan guided reaction. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6).

[Ruspini and Ruspini, 1991] Ruspini, E. H. and Ruspini, D. 1991. Autonomous vehicle motion planning using fuzzy logic. In *Procs. of the IEEE Round Table on Fuzzy and Neural Sys. and Vehicle Appl.*, Tokyo, Japan.

[Ruspini, 1990] Ruspini, E. H. 1990. Fuzzy logic in the Flakey robot. In *Procs. of the Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA)*, Japan. 767-770.

[Ruspini, 1991a] Ruspini, E. H. 1991a. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning* 5.

[Ruspini, 1991b] Ruspini, E. H. 1991b. Truth as utility: A conceptual systesis. In *Procs. of the 7th Conf. on Uncertainty in Artificial Intelligence*, Los Angeles, CA.

[Saffiotti *et al.*, 1993] Saffiotti, A.; Konolige, K.; and Ruspini, E. H. 1993. Now, do it! Technical report, SRI Artificial Intelligence Center, Menlo Park, California. Forthcoming.

[Saffiotti, 1993] Saffiotti, A. 1993. Some notes on the integration of planning and reactivity in autonomous mobile robots. In *Procs. of the AAAI Spring Symposium on Foundations of Automatic Planning*, Stanford, CA.

[Sugeno and Nishida, 1985] Sugeno, M. and Nishida, M. 1985. Fuzzy control of a model car. *Fuzzy Sets and Systems* 16:103-113.

[Yen and Pfluger, 1992] Yen, J. and Pfluger, N. 1992. A fuzzy logic based robot navigation system. In *Procs. of the AAAI Fall Symposium on Mobile Robot Navigation*, Boston, MA. 195-199.

[Zadeh, 1978] Zadeh, L.A. 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1:3-28.

CARMEL vs. *Flakey*: A Comparison of Two Winners

CARMEL Versus FLAKEY
A Comparison of Two Winners

Clare Congdon, Marcus Huber, David Kortenkamp, Kurt Konolige, Karen Myers, Alessandro Saffiotti, and Enrique H. Ruspini[1]

Reprinted from

AI magazine

Volume 14, Number 1 (Spring 1993)

CARMEL Versus FLAKEY A Comparison of Two Winners

Clare Congdon, Marcus Huber, David Kortenkamp, Kurt Konolige, Karen Myers, Alessandro Saffiotti, and Enrique H. Ruspini[1]

■ The University of Michigan's CARMEL and SRI International's FLAKEY were the first- and second-place finishers, respectively, at the 1992 Robot Competition sponsored by the American Association for Artificial Intelligence. The two teams used vastly different approaches in the design of their robots. Many of these differences were for technical reasons, although time constraints, financial resources, and long-term research objectives also played a part. This article gives a technical comparison of CARMEL and FLAKEY, focusing on design issues that were not directly reflected in the scoring criteria.

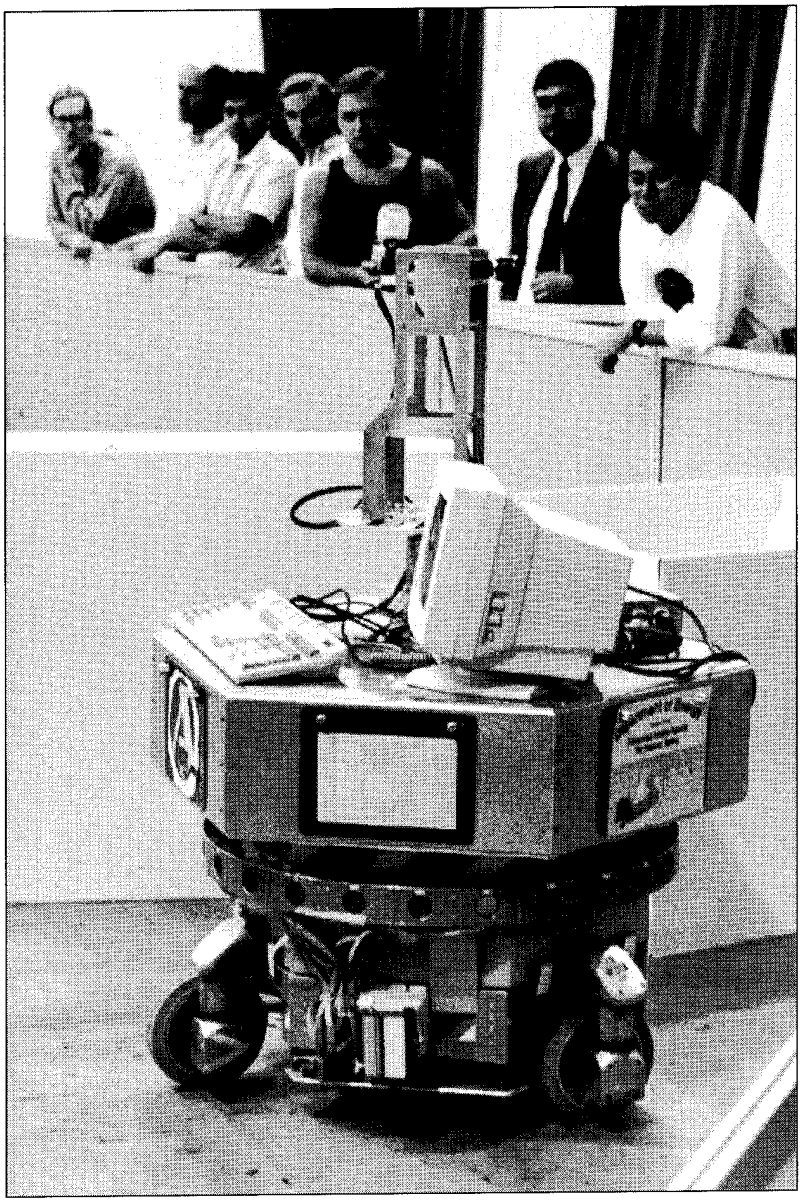
The University of Michigan's CARMEL and SRI International's FLAKEY were the first- and second-place finishers, respectively, at the 1992 Robot Competition sponsored by the American Association for Artificial Intelligence (AAAI) (see the Dean and Bonasso article in this issue). Interestingly, the two teams used vastly different approaches in the design of their robots. Many of these differences were for technical reasons, although time constraints, financial resources, and long-term research objectives also played a part.

The final scores for the robots, based solely on competition-day performance, constitute only a rough evaluation of the merits of the various systems. This article provides a technical comparison of CARMEL and FLAKEY, focusing on design issues that were not directly reflected in the scoring criteria. Space limitations preclude detailed descriptions of the two approaches; further details can be found in an upcoming AAAI Technical Report by the authors.

The Two Robots

CARMEL (computer-aided robotics for maintenance, emergency, and life support) is based on a commercially available Cybermotion K2A mobile robot platform. CARMEL is a cylindrical robot about a meter in diameter, standing a bit less than a meter high. It has a top velocity of 780 millimeters/second and a top turning rate of 120 degrees/second; it moves using three synchronously driven wheels. For sensing, CARMEL has a ring of 24 Polaroid sensors and a single black-and-white charge coupled device camera. The camera is mounted on a rotating table that allows it to turn 360 degrees independently of robot motion. Three computers work cooperatively while the robot is running: First, an IBM PC clone runs a 33-MHz, 80486-based processor that performs all top-level functions and contains a frame grabber for vision processing. Second, a motor-control processor (Z80) controls the robot's wheel speed and direction. Third, an IBM PC XT clone is dedicated to the sonar ring. All processing and power are contained on board CARMEL.

CARMEL's software design is hierarchical in structure. At the top level is a supervising planning system that decides when to call subordinate modules for movement, vision, or the recalibration of the robot's position. Each of the subordinate modules is responsible for doing low-level error handling and must return control to the planner in a set period of time, perhaps reporting failure; the planning module then determines whether to recall the submodule with different parameters or resort to another course of action.



CARMEL avoids obstacles using a point-to-point, goal-directed algorithm called VFH.

CARMEL avoids obstacles using a point-to-point, goal-directed algorithm called VFH (Borenstein and Koren 1991a, 1991b). Object recognition is done using a single camera and a one-pass algorithm to detect horizontally striped, bar-code–like tags on each of the 10 objects. A distance and a heading for each object are returned. Recalibrating the robot's position is done by triangulating from three objects with known locations.

The software system of CARMEL was kept modular to allow for a team design, whereby small groups could work independently on each module. Using this approach, the team of 30 students was able to write its winning system in six months. Only the low-level object-avoidance modules existed before work on the competition began. CARMEL's software system was also kept simple so that it could be run completely on board, allowing CARMEL to navigate at high speeds while it smoothly avoided obstacles. (Many of the other robots in the competition were sending information to off-board processors and, as a result, operated in a jerky, stop-and-go fashion, moving a bit but then having to stop and wait while sensor information was sent off board and processed and the results transmitted back to the robot.)

FLAKEY is a custom-built mobile robot platform approximately 1 meter high and .6 meter in diameter. There are two independently driven wheels, 1 on each side, giving a maximum linear velocity of 500 millimeters/second and a turning velocity of 100 degrees/second. Like CARMEL, FLAKEY has ultrasonic sonar sensors good to about 2 meters, but instead of a uniform ring, FLAKEY has 4 sensors facing front, 4 facing back, and 2 facing each side. Additionally, FLAKEY has 8 touch-sensitive bumpers around the bottom perimeter of the robot and a structured-light sensor that is a combination of a light stripe and a video camera that is capable of providing a dense depth map over a small area in front of FLAKEY. FLAKEY has 3 computers: (1) a Z80 motor and sonar controller, (2) a SUN 3 dedicated to the structured-light sensor, and (3) a SPARCSTATION responsible for high-level routines. During the competition, all computation was performed on board.

FLAKEY's basic software design is distributed: The modules work in parallel and communicate through a blackboardlike structure called the *local perceptual space* (LPS). LPS is a geometric egocentric map of the area within two meters of the robot. Modules contribute information to, and draw information from, LPS. The loosely linked structure makes it

possible to have tasks running in parallel that have different reaction-time and information requirements. On the perception side, modules add raw sonar and structured-light information to LPS, treating it as an occupancy grid. Other interpretive processes use this information to construct and maintain higher-order structures, parsing the data into surface segments, recognizing objects, and so on. All this information is coordinated geometrically so that an action module can use whatever form is appropriate, for example, the occupancy grid for obstacle avoidance, surface segments for path planning, and object tags for task planning.

On the action side, there are three main types of modules. At the lowest level, reactive-action modules called *behaviors* guide the robot's movements. The input to these modules is the occupancy grid for obstacle avoidance plus artifacts (such as a path to follow) that are put into LPS by higher-level navigation routines. FLAKEY was unique in using fuzzy rules as the building block for behaviors (Saffiotti and Ruspini 1993), giving it the ability to react gracefully to the environment by grading the strength of the reaction (for example, turn left) according to the strength of the stimulus (for example, the distance of an obstacle on the right).

More complex behaviors, such as moving to desired locations, use surface information and artifacts to guide the reactive behaviors; they can also add artifacts to LPS as control points for motion. At this level, fuzzy rules allow FLAKEY to blend possibly conflicting aims into one smooth action sequence. At a higher level, the navigation module autonomously updates FLAKEY's global position by comparing it to a tolerant global map, which contains prior, approximate spatial knowledge of objects in the domain. Finally, task-level modules continuously monitor the progress of the complex behaviors, using information from the navigation module to plan sequences of behaviors to achieve a given goal.

The distributed architecture and loosely coupled control structure enable FLAKEY to simultaneously interpret sensory data, react to the local environment, and form long-range plans. The modular and distributed design of FLAKEY means that it is both flexible and extensible. The SRI team incorporated large portions of software previously written for an office environment, including almost all the perceptual routines and the low-level behaviors. The team started working on the competition one month before it began and

The distributed architecture and loosely coupled control structure enable FLAKEY to simultaneously interpret sensory data, react to the local environment, and form long-range plan.

CARMEL was distinguished by its graceful motion around obstacles in open terrain and was pleasant to watch

produced and integrated modules for complex behaviors, tasks, and navigation.

Issues in Moving

In stage 1, both robots had to roam the arena, avoiding people and obstacles. Both robots used sonar sensors as their primary obstacle-avoidance sensors. CARMEL used a two-step process in which the sonar readings were first filtered to reduce noise and imprecision (Borenstein and Koren 1992), and then an occupancy grid–style map, similar to that introduced by Moravec and Elfes (1985), was updated. CARMEL used this sonar map to navigate. Similarly, FLAKEY used LPS to integrate sonar readings and fuzzy-control rules to control motion.

Both robots performed obstacle avoidance remarkably well despite attempts by the judges to surprise and contain them. FLAKEY's use of fuzzy rules resulted in extremely smooth and reliable movement. FLAKEY uses two-part obstacle-avoidance rules: Longer-range rules deflect FLAKEY away from distant obstacles, and collision-avoidance rules force emergency maneuvers when an object is suddenly detected nearby. These rules are typically combined with rules for purposeful motion, such as following a wall. This combination of rules ensures that strategic goals are achieved as much as possible while a high reactivity is maintained. In this phase of the competition, speed was limited to 200 millimeters/second, primarily because of the blind spots on the diagonal: Objects in these positions had to be relatively close before they could be seen by the sonar. FLAKEY's reliable behavior is best summarized by one judge's comment: "only robot I felt I could sit or lie down in front of" (which he actually did).

CARMEL was distinguished by its graceful motion around obstacles in open terrain and was pleasant to watch. It moved at a speed of 300 millimeters/second, noticeably faster than FLAKEY. However, under prodding from the judges, CARMEL touched two obstacles and grazed a judge. CARMEL touched objects in part because many variables in CARMEL's obstacle-avoidance code need to be tuned for the environment in which it is running. The Michigan team had assumed an environment with dynamic but much more benign obstacles.

FLAKEY placed ahead of CARMEL in this stage of the competition and was only one point behind the first-place (at this point) entry, TJ2 from IBM. Part of the reason why CARMEL did not do as well was because it was so goal oriented; that is, it was always trying to get somewhere in particular. CARMEL could not be "shepherded" about by the judges because it had a dogged persistence in trying to achieve its goal location. Both teams noticed that behavior could be improved markedly by tuning the parameters of the avoidance routines.

Issues in Object Recognition

Stages 2 and 3 both required the ability to detect and visit objects (specifically, poles of a fixed diameter) scattered throughout the arena. The rules permitted teams to modify poles to facilitate the recognition process, although a small bonus was awarded for full autonomy, that is, no altering of the environment. Michigan took advantage of the object-modification rule by attaching a distinct omnidirectional bar-code tag to each pole. CARMEL's vision algorithm was designed to extract the bar codes from an image.

SRI, however, was one of only two teams (the other being Brown University) that did not modify the arena or poles in any way. Instead, FLAKEY used only the physical characteristics of the poles themselves in the detection process. FLAKEY used a two-tiered approach, whereby sonar input was monitored during navigation to detect candidate poles, and candidates were actively verified by having FLAKEY navigate to a position where the structured-light sensor could be applied. This hybrid approach was necessary because of the limitations of the two sensing modalities: Structured-light verification is highly accurate but applies only to a small perceptual space (less than two meters) directly in front of the robot; sonar input covers a much larger space during navigation but is not nearly as reliable for object recognition. Both the structured-light and sonar routines were built using low-level perceptual routines that FLAKEY has used for some time.

The recognition components of both teams performed extremely well during the competition. CARMEL never saw a false object, and it never missed seeing an actual object; similarly, FLAKEY's structured-light routine was perfectly reliable. CARMEL's performance was surprising because its long-range vision created the added difficulty of dealing with false objects outside the arena, a problem that FLAKEY's short-range sensors did not have. FLAKEY's candidate generation techniques based on sonar input also performed well, picking out only two nonpoles (box corners)

as candidates and only failing to detect one pole in its perceptual space as a candidate (because the robot passed too close to the pole).

The SRI team demonstrated that reliable object-type recognition was possible using only physical characteristics of the objects and simple domain constraints (such as nonproximity to other objects) without having to modify the environment. As such, FLAKEY, in contrast to CARMEL, was able to perform recognition for classes of objects rather than specially marked individuals in the class. One consequence of doing class recognition was that individual objects had to be identified based solely on information about the object's location. In contrast, the individualized bar codes used by the Michigan team provided immediate identification information to CARMEL.

CARMEL's use of long-range sensing made it possible to locate objects from as far away as 12 meters (over half the diameter of the arena). In contrast, FLAKEY could only recognize poles and candidates within its local perceptual space. As discussed later, this difference had a major impact on the methods used by the two teams for mapping and navigation.

Issues in Mapping

To be competitive in stage 3, it was necessary for the robots to generate maps of the environment during stage 2. At a minimum, these maps contain the location of discovered poles, but they could also encode further information, such as the positions of obstacles or walls. A complementary problem to map construction is *self-localization*, which involves having the robot determine where it is relative to the map. A critical issue faced by both robots in solving these problems was the inaccuracies inherent in *dead reckoning*, the robot's internal calculation of its location based on wheel movements.

Automated map generation remains a topic of current research for the field of robotics. Michigan and SRI chose two different approaches to the design of maps for their robots. CARMEL used a global Cartesian system that stores only pole locations and the current position of the robot. To track its position with reference to the map, CARMEL relied exclusively on dead reckoning: When initialized, it was given its position and orientation on the map, and subsequent movements gave an estimated position based on wheel rotation. When discovered, the poles were placed on the map using the current estimated position together with the range and angle returned by the vision system.

Of course, errors in estimated position accumulate over time from wheel slippage and the like; CARMEL incorporated an algorithm to triangulate its position from known object locations, thus reducing the error. The vision-based triangulation system was not actually used for the competition because of last-minute changes to the system software. However, not using triangulation did not noticeably affect the performance of CARMEL for three reasons: First, the time and the distance between tasks were small; second, CARMEL's dead reckoning and its vision system were highly accurate; and, finally, the planning system was designed to deal with self-localization errors. CARMEL could be several meters away from the expected location of the pole and still be able to locate it.

In contrast to CARMEL, FLAKEY used a tolerant global map containing local Cartesian patches related by approximate metric information. Each patch contains some recognizable feature or landmark by which the robot can orient itself; the approach is similar to the work on landmark-based navigation (Kuipers 1978). The patches chosen for the competition were the walls of the arena because they were the most stable features for navigation. The approximate length and the relative orientation of the walls were given to FLAKEY as prior knowledge; FLAKEY could easily have learned this information by circumnavigating the arena.

The SRI team chose the tolerant global maps because FLAKEY accumulated dead-reckoning errors more quickly than CARMEL. Moving four or five meters, especially with some turning, can cause significant errors in estimated position, and FLAKEY must use sensed landmarks to correct its localization on the map. Compounding the problem is FLAKEY's short-range sensing, which makes it impossible to locate landmarks more than a few meters away. The tolerant global maps are a solution to FLAKEY's imprecision in large-scale sensing and movement. Within each patch, local landmarks can be sensed almost continuously (in this case, the arena walls and wall junctions) to keep localized. When going between patches, approximate metric information can be used to find the next landmark for localization. Because sensing and movement are accurate only over small distances, there is no need to keep a highly precise global geometry; further, FLAKEY would find it impossible to use this information.

Fuzzy rules allow FLAKEY to blend possibly conflicting aims into one smooth action sequence

The recognition components of both teams performed extremely well

The use of precise dead reckoning and long-range sensing gave CARMEL a marked advantage over FLAKEY in the competition because it made it easy to both register the poles in a global coordinate system and determine trajectories for navigating from one to another (as required for stage 3). FLAKEY's ability to rerecognize poles that it discovered previously demonstrated that the tolerant global map can successfully be used for self-localization, although FLAKEY encountered some difficulties in using this system (see the next section).

It is interesting to speculate about how well the different methods would work in other domains. FLAKEY's tolerant global maps were designed for an office environment, where navigation landmarks are plentiful (walls, corridors, doors), and long-range triangulation is difficult and of limited value. The tolerant global maps are robust in this situation, whereas a precise global Cartesian map would be hard to acquire and use. Its main advantage—navigation over open space—would be minimized because most navigation is by corridor paths.

However, FLAKEY's approach is less useful in more open areas such as outdoor navigation, where paths and local landmarks might be sparse. In this case, CARMEL benefits from the inclusion of more global positioning information.

Issues in Navigation

Task-oriented navigation played a critical role in the competition. Stage 2 required explorative navigation of the arena to detect and visit poles. Because the robots had no prior information about object locations, a general and thorough exploration methodology was required. Stage 3 involved directed navigation: Robots were to revisit three poles in a prespecified sequence and then return home.

Explorative Navigation

CARMEL's long-range object-recognition capabilities enabled the Michigan team to use a fairly simple exploration strategy. CARMEL's exploration consisted of moving to viewing positions distributed throughout the arena, executing a visual sweep for objects, and then visiting each object.

FLAKEY's reliance on local sensing necessitated an actual physical exploration of the environment: To ensure full coverage, FLAKEY had to cover the full extent of the competition arena with its local perception. The strategy adopted by the SRI team was to traverse the perimeter of the arena, making forays into the center of the arena at certain points along the way. This strategy was designed to reconcile the conflicting objectives of providing broad coverage of the arena and keeping FLAKEY self-localized using information about wall locations.

FLAKEY did encounter some localization problems near the end of its stage 2 run, primarily because of a tactical mistake (on the part of the designers!) in the execution of forays. FLAKEY initiated its final foray before having registered the current wall. As a result, dead-reckoning errors accumulated to such an extent that FLAKEY's beliefs about its position were inaccurate. Given more time, FLAKEY would eventually have returned to the wall and reregistered itself, thus correcting the problem. The entire issue could have been avoided had forays been postponed until wall registration had taken place.

The Michigan team's use of long-range sensing easily enabled CARMEL to find all 10 poles within the allotted 20-minute search period. In contrast, the physical exploration executed by FLAKEY was time consuming. In the end, FLAKEY found and correctly registered only 8 of the 10 poles before time expired. Certainly, the Michigan approach was superior given the conditions of the competition environment. In particular, Michigan took full advantage of the fact that objects would be visible above all obstacles. Because FLAKEY's method was not based on any such assumptions, it was less efficient; however, FLAKEY's exploration method could be used in more realistic environments, where objects can be occluded.

Directed Navigation

In stage 3 of the competition, the robots were given three poles to visit in order, and then, they were to return to a home position. This stage was timed, with the robots receiving points based on their time with respect to the other robots.

FLAKEY's strategy of registering objects and itself with respect to walls meant that the robot had to navigate along the perimeter of the arena when traveling between objects. Visiting an object registered with respect to a wall W involved determining the direction of the shortest perimeter path to W (either clockwise or counterclockwise), following the perimeter in this direction until W was encountered, and then using dead reckoning within the coordinate system of W to move to the pole. CARMEL, however, used dead reckoning with its global map to proceed directly

to the recorded locations of objects. Not surprisingly, CARMEL was able to perform the stage-3 visiting task in a much shorter period of time (3 minutes versus 11 minutes for FLAKEY).

Like many other teams, the Michigan team marked the home position by placing an 11th pole there. This modification to the environment was made to provide a perceptual landmark for the home position, which compensated for the accumulation of dead-reckoning errors during the run. FLAKEY, in contrast, did not require any such modification to the environment. Instead, it was able to treat the home location in the same manner as other positions of interest (such as pole locations or foray positions) because it used the tolerant global maps to continuously correct its position.

Conclusion

Both CARMEL and FLAKEY must be considered unqualified successes, having bested 10 or so other entries in a nationwide competition. There were many reasons for this success. Both teams did all their processing on board the robot, avoiding problems with radio and video links and enabling their robots to be more reactive. Both teams inherited robots that had well-developed software systems. In addition, both teams used simulations to speed the development process. However, the approach to the competition was different for each team. Michigan looked at the competition rules and engineered a system to optimize its robot's performance at the cost of generality. SRI used the competition as a demonstration of the application of its research in a new domain, without engineering any hardware specific for this domain.

Interestingly, neither team used any geometric planning for navigation around obstacles to a goal point, although this area is a large part of robotics research (Latombe, Lazanas, and Shekhar 1991). Instead, both teams relied on the simple strategy of heading toward the goal and using reactive behavior to avoid obstacles, with simple methods for getting out of cul-de-sac situations. Geometric planning requires some sophistication in perception and mapping of obstacles and can be difficult to perform in real time. The large openings around obstacles in the competition made it easy to pursue simpler strategies, and we speculate that in other domains, geometric planning will also play a minor role in navigation.

It is interesting to try to compare the two system architectures. At the level of reactive movement, FLAKEY perhaps had the advantage, because the fuzzy-control paradigm provides a flexible and powerful representation for specifying behavior. In less than a month, the SRI team was able to write and debug half a dozen complex movement routines that integrated perception and action in the service of multiple simultaneous goals.

In terms of overall design, it is difficult to compare the relative merit of the two architectures because the approaches to solving the problem were so different. FLAKEY's distributed control scheme allows various modules to run in parallel, so that (for example) self-localization with respect to landmarks occurs continuously as FLAKEY moves toward a goal location or searches for poles. However, the distributed design leads to behavior that is more difficult to predict and debug than that of CARMEL's top-down approach in which all the perception and goal actions are under sequential, hierarchical control.

Although Michigan was the winner of the competition, it is not clear that its system can easily be extended to other domains. Certainly, the obstacle-avoidance routines are necessary in any domain and are widely applicable. CARMEL's reliance on a global coordinate system and tagged objects restricts it to engineered environments that can accurately be surveyed (a reasonable assumption when you consider how much of the world in which humans operate is highly engineered). Also, CARMEL's simple exploration strategies would be naive in an environment where objects can be occluded. CARMEL's keys to victory were fast, graceful obstacle avoidance and fast, accurate vision algorithms, not cognitive smarts.

FLAKEY, moving more slowly and possessing less accurate and more local sensing, had to rely on a smart exploration strategy and constant position correction. One of the key research ideas behind FLAKEY is that natural (that is, non-engineered) landmarks are sufficient if the right map representation is used, and it was gratifying to see this approach work in a new environment. Still, FLAKEY could be more efficient in navigating open spaces if it incorporated more global geometric information, such as CARMEL used.

The fact that CARMEL, which is sensor rich and cognitively poor and FLAKEY, which is sensor poor and cognitively rich, came in as the top two robots in the competition clearly shows that fundamental trade-offs can be made in engineering mobile robots. Complex sensing can allow for simple planning; simple

sensing requires complex planning. In no sense is either robot more complex than the other; it is just that the complexity lies in different places. What was not clear from the competition was whether complex sensing and complex planning will make for a fundamentally better robot. This issue will have to be resolved at future competitions.

Acknowledgments

Support for several of the Michigan authors was from Defense Advanced Research Projects Agency grant DAAE-07-92-C-R012. Purchase of CARMEL and its computers, development of the EERUF (error elimination rapid ultrasonic firing) and VFH (vector field histogram) algorithms, and support of several Michigan authors were through Department of Energy grant DE-FG02-86NE37969. Additional financial support for the Michigan team was from The University of Michigan College of Engineering, The University of Michigan Rackham Graduate School, ABB/Graco Robotics Inc., and the American Association for Artificial Intelligence.

Support for Kurt Konolige and Karen Myers came partially from Office of Naval Research contract N00014-89-C-0095. Alessandro Saffiotti is supported by a grant from the Italian National Research Council. We gratefully acknowledge the help of Enrique Ruspini and Nicolas Helft on the fuzzy-control system and Daniela Musto in suggesting ideas for self-localization and goal-directed movement.

Note

1. Clare Congdon, Marcus Huber, and David Kortenkamp are with the AI Laboratory at the University of Michigan, and Kurt Konolige, Karen Myers, Alessandro Saffiotti, and Enrique Ruspini are with SRI International.

References

Borenstein, J., and Koren, Y. 1992. Noise Rejection for Ultrasonic Sensors in Mobile Robot Applications. In Proceedings of the IEEE Conference on Robotics and Automation, 1727–1732. Washington, D.C.: IEEE Computer Society.

Borenstein, J., and Koren, Y. 1991a. Histogramic In-Motion Mapping for Mobile Robot Obstacle Avoidance. *IEEE Transactions on Robotics and Automation* 7(4): 535–539.

Borenstein, J., and Koren, Y. 1991b. The Vector Field Histogram for Fast Obstacle Avoidance for Mobile Robots. *IEEE Transactions on Robotics and Automation* 7(3): 278–288.

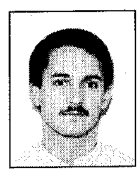
Kuipers, B. 1978. Modeling Spatial Knowledge. *Cognitive Science* 2:129–153.

Latombe, J-C.; Lazanas, A.; and Shekhar, S. 1991. Robot Motion Planning with Uncertainty in Control and Sensing. *Artificial Intelligence* 52(1): 1–47.

Moravec, H., and Elfes, A. 1985. High-Resolution Sonar Maps from Wide-Angle Sonar. In Proceedings of the IEEE Conference on Robotics and Automation, 19–24. Washington, D.C.: IEEE Computer Society.

Saffiotti, A., and Ruspini, E. H. 1993. Blending Reactivity and Goal Directedness in a Fuzzy Controller. In Proceedings of the Second International IEEE Conference on Fuzzy Systems. Washington, D.C.: IEEE Computer Society. Forthcoming.

Clare Bates Congdon is a doctoral candidate in the AI Laboratory of the Electrical Engineering and Computer Science Department at the University of Michigan. She holds a B.A. in mathematics from Wesleyan University and an M.S. in computer science from the University of Michigan. Her thesis work is a comparison of genetic algorithms and other machine learning approaches in a complex domain characterized by nonlinear interactions among features (the genetic and biochemical traits that predict coronary artery disease).



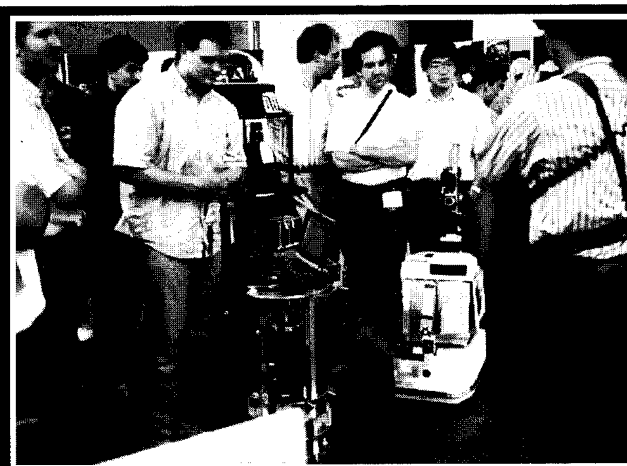
Marcus Huber is a research assistant at the University of Michigan AI Laboratory. He graduated with a B.S. in electrical engineering from GMI Engineering and Management Institute in 1988 and received his M.S. in computer science from the University of Michigan in 1991. His research interests include distributed AI, plan recognition, and autonomous robotics. He is a member of the Institute of Electrical and Electronics Engineers, the Association of Computing Machinery, and the American Association for Artificial Intelligence.

David Kortenkamp is a Ph.D. candidate at the University of Michigan AI Laboratory. He graduated with a B.S. in computer science from the University of Minnesota in 1988 and received his M.S. in computer science from the University of Michigan in 1990. His thesis research involves applying principles of cognitive mapping to mobile robot navigation tasks.

Kurt Konolige is a senior computer scientist in the AI Center at SRI International. He received his Ph.D. in computer science from Stanford University in 1984; his thesis, "A Deduction Model of Belief and Its Logics," developed a model of belief based on the resource-bounded inferential capabilities of agents. His research interests are broadly based on issues of commonsense reasoning, including introspective reasoning, defeasible reasoning, and reasoning about cognitive state. More recently, he has

AAAI-93 Robot Exhibition

July 13–15, 1993
Washington, DC



Following the highly successful robotics exhibition at AAAI-92, AAAI is planning to hold a robot competition at the national conference in Washington D.C. in July of 1993

Last year's competition was a three-stage event in which mobile robots demonstrated skills of reactivity, exploration, and directed search (a detailed description is in the Summer 1992 issue of *AI Magazine*).

Mobile robotics is an area where much of the research in diverse AI areas can be effectively and creatively combined to give interesting results. At AAAI-93, we would like to extend the competition to highlight as wide a range of robotic research as possible, and to stress the "intelligent" aspects of their behavior. In addition to mobile robots, we are also considering having a competition among robotic manipulators, either stationary or attached to mobile platforms.

If you are interested in more detailed information about the competition, please contact:

Kurt Konolige, Artificial Intelligence Center
SRI International, 333 Ravenswood Avenue
Menlo Park, CA 94025, (konolige@ai.sri.com)

or

Reid Simmons, School of Computer Science
Carnegie Mellon University,
5000 Forbes Avenue
Pittsburgh, PA 15213,
(reids@cs.cmu.edu)

been working on making moving robots more intelligent.

Karen Myers is a computer scientist in the AI Center at SRI International. She completed her Ph.D. in computer science at Stanford University in 1991. Her research interests include the acquisition and use of maps by mobile robots, reactive planning, and hybrid reasoning systems.

Alessandro Saffiotti is a research assistant at the Université Libre de Bruxelles, Belgium. He received his Ph.D. in computer science from the University of Pisa, Italy, in 1988. Since January 1992, he has been at SRI International as an international fellow, working on the use of multivalued logics in the control of autonomous mobile robots. His research interests include the formalization of partial belief and uncertain knowledge and their use in reasoning, planning, and acting.

Enrique H. Ruspini is a senior computer scientist with the AI Center of SRI International. He received his doctoral degree from the University of California at Los Angeles in 1977. He has done extensive research on fuzzy sets and their applications, having pioneered the application to classification problems in 1969.

Robust Execution of Robot Plans using Fuzzy Logic

IJCAI-93 Workshop on Fuzzy Logic in Artificial Intelligence
Chamberry, France, August 1993

Robust Execution of Robot Plans using Fuzzy Logic

Alessandro Saffiotti* **Enrique H. Ruspini** **Kurt Konolige**

Artificial Intelligence Center
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025, U.S.A.
{saffiotti, ruspini, konolige}@ai.sri.com

1 Introduction

Performance of complex tasks by an autonomous robot requires careful planning. A large part of AI research has been devoted to the study and development of general knowledge-based planning techniques [Wilkins, 1988]. In order to make the planning problem tractable, most of these techniques make use of strong simplifying assumptions about the conditions under which the plan is to be executed. Typically, the environment is supposed to have simple, well-known, predictable dynamics — e.g., to be static; agent's knowledge about this environment is supposed to be complete and accurate; and agent's actions are supposed to be completely reliable. Unfortunately, in any realistic case involving an autonomous mobile robot's operation in the real world, none of these assumptions holds. Prior information is in general approximate and incomplete; sensorial input is noisy and limited by the sensors' range and environmental the features (e.g., occlusion); the dynamics of the environment is largely unpredictable (e.g., people moving around); and robot's actions are imprecise and may fail. This has led many researchers to question the validity of traditional AI planning techniques for situated agents [Brooks, 1987], and to propose alternative (or in some cases complementary) approaches (e.g., [Firby, 1987; Schoppers, 1987; Saffiotti, 1993]).

In this paper, we propose an approach to plan execution that can cope with many of the difficulties encountered in real-world situations. We exploit the flexibility of fuzzy logic for dealing with the imprecision and errors in the prior knowledge, in the sensed information, and in robot's movement. We describe a fuzzy controller that implements robust high-level actions (like traversing a hallway, or crossing a door), and provides capabilities for:

*On leave from Iridia, Université Libre de Bruxelles, 50 av. F. Roosevelt, Brussels, Belgium.

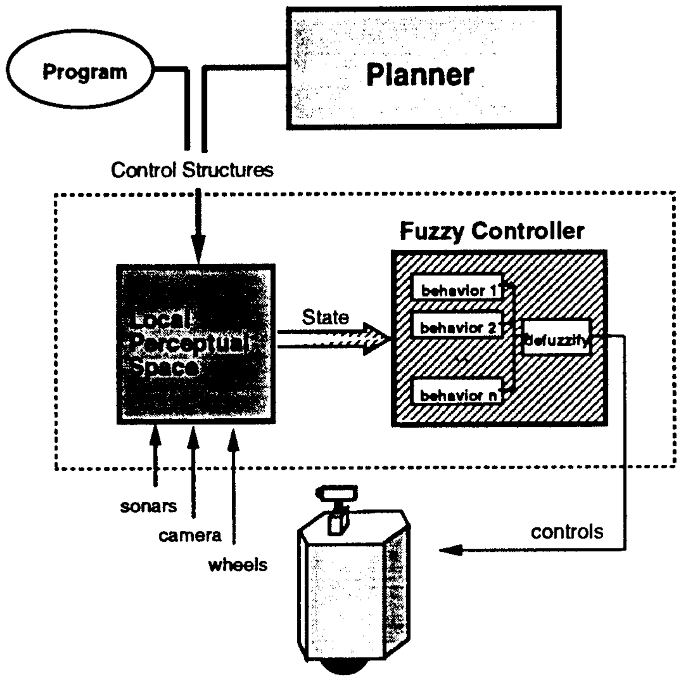


Figure 1: Flakey's architecture (partial).

1. Robust, uncertainty-tolerating goal-directed activity;
2. Real-time reactivity to unexpected contingencies (e.g., unknown obstacles); and
3. Blending of multiple goals (e.g., reaching a given location while avoiding obstacles and people).

goals are represented by *control structures*: abstract descriptions of actions that can be easily generated by traditional AI planning techniques, and can be used by the fuzzy controller to orient the robot's activity [Saffiotti *et al.*, 1993a].

The proposed approach has been implemented on our mobile robot, Flakey, and consistently validated in an unmodified office environment during normal office activity. Flakey's fuzzy controller has also been demonstrated at the first AAAI robot competition, where Flakey placed second, and was prized for its smooth reliable reactivity [Congdon *et al.*, 1993]. A full account of the fuzzy controller can be found in [Saffiotti *et al.*, 1993b].

Figure 1 sketches the architecture of Flakey. The controller is centered on the notion of *behavior*. Intuitively, a behavior is one particular control regime that focuses on achieving one specific, predetermined goal, like avoiding obstacles, or reaching a specific location

(given as a parameter). Behaviors take their input from a common storage structure, named *local perceptual space* (LPS), where all sensory input (possibly after some perceptual interpretation) is maintained. Moreover, the LPS may contain *control structures*, representing strategic goals to achieve, like a position to reach or a door to cross. Specific behaviors in the fuzzy controller react to the presence of control structures in the LPS by trying to satisfy them. Control structures can be put in the LPS by a human programmer, or by an automatic planner.

Two important aspects differentiate this architecture from other two-level architectures [Hanks and Firby, 1990]. First, most current approaches typically assume that behaviors are implemented by executable procedures that directly control robot's effectors, and arbitration between competing behaviors results in giving complete contol to one of them. By contrast, our elementary behaviors only express *preferences* among possible control actions. Many behaviors are in general active at the same time: the fuzzy controller weights the preferences expressed by all the active behaviors, and generate one tradeoff control for the effectors.

The second difference is that in our approach plans do not constitute *programs* to be executed by the controller, but *information* about which goals should be considered, and when. (This is why control structures are put into the LPS, rather than being directly fed to the controller). The fuzzy controller exploits this information, together with the perceptual data, in order to take decisions about the best actions to perform. Viewing plans as sources of information rather than as sequences of instructions provides added flexibility in face of unknown execution contingencies [Suchman, 1987; Schoppers, 1987; Payton *et al.*, 1990].

2 The fuzzy controller

The basic building block of the fuzzy controller is a *behavior*. A behavior implements a motor skill of the agent, aimed at achieving a given goal. Behavioral skills are expressed as *preferences* over possible control actions from the perspective of achieving that behavior's goal. For example, a behavior aimed at following a given wall could prefer actions that keep the agent parallel to that wall at a "safe" distance.

More formally, and following the semantic characterization of [Ruspini, 1991a; Ruspini, 1991b], we describe each behavior B in terms of a desirability function

$$Des_B : \text{State} \times \text{Control} \rightarrow [0,1]$$

that measures, for each state s (i.e., input configuration in the LPS) and control vector[1] c, the desirability $Des_B(s,c)$ of applying control c in the state s *from the point of view of* B. Hence, a behavior maps each input configuration to a fuzzy set of admissible controls — this contrasts with other approaches to robot control where individual behaviors map each state to *one* preferred control (e.g., [Khatib, 1986; Brooks, 1987; Arkin, 1990; Gat, 1991]).

[1]In the case of Flakey, control vectors include linear acceleration and turning angle.

For any behavior B, the task of our fuzzy controller is to compute, at every step, the value of $Des_B(s, c)$ from the current LPS configuration s, and then choose one control $\hat{c}$ for actual execution. This choice corresponds to what is commonly referred to as "defuzzification" in the fuzzy control literature. We currently use the so called centroid defuzzification:

$$\hat{c} = \frac{\int c\, Des_B(s, c)\, dc}{\int Des_B(s, c)\, dc}. \tag{1}$$

In practice, we approximate desirability functions by sets of fuzzy rules of the form

IF A_i THEN C_i

where A_i is composed of fuzzy predicates and fuzzy connectives,[2] and C_i is a fuzzy set of control vectors. For example, the following rule is part of the rules implementing the KEEP-OFF behavior, a behavior intended to keep Flakey safely away from unknown obstacles as they are perceived by the sonars:

```
IF obstacle-close-in-front
AND NOT obstacle-close-on-left
THEN turn sharp-left
```

Given a ruleset $\mathcal{R} = \{R_1, \ldots, R_n\}$, and a state s, the fuzzy controller first computes

$$Des_{\mathcal{R}}(s, c) = (A_1(s) \circledast C_1(c)) \oplus \ldots \oplus (A_n(s) \circledast C_1(c)) \tag{2}$$

and then chooses one control $\hat{c}$ to apply, using the (1) above. It is the task of the programmer to make sure that $Des_{\mathcal{R}}$ is a "reasonable" approximation of the intended desirability function Des_B — i.e., that the rules produce the expected effect.[3] In practice, we have found that a limited number of rules is sufficient to achieve a good performance of each behavior. For instance, the KEEP-OFF behavior above consists of four rules; in cluttered spaces, the combination of these rules through (2) has been shown to actually produce effective maneuvers towards open areas.

The controller includes "purposeful" behaviors that take explicit goals into consideration, represented by control structures. Purposeful behaviors are also described by desirability functions of the form $Des_B(s, c)$, the only (formally invisible) difference being that the state s —and hence the antecedents A_i's of the fuzzy rules— depends in general on some property of an *artifact* (see below) in the LPS. For instance, the FOLLOW-WALL behavior includes the following rule:

```
IF wall-too-far-on-right
THEN turn moderate-right
```

[2] We use min, max and complement to 1 for $\circledast$ (AND), $\oplus$ (OR), and $\oslash$ (NOT).

[3] We are currently exploring formal techniques to automatically generate correct rules from abstract specifications of goals [Ruspini and Saffiotti, 1993].

Many behaviors can be simultaneously active in the fuzzy controller, each aimed at one specific goal (see Figure 1). The fuzzy controller selects the controls that best satisfy all the active behaviors. However, not all behaviors are always applicable: for instance, the FOLLOW-WALL behavior is most applicable when the wall is near and the path is clear; while KEEP-OFF becomes more applicable when there is an obstacle on the way. To account for this, we associate with each behavior B a *context of applicability*, expressed by a fuzzy predicate Cxt_B. Given n behaviors $\{B_1, \ldots, B_n\}$, the fuzzy controller combines their desirability functions, modulo their contexts, into one overall desirability function

$$Des(s,c) = (Des_1(s,c) \circledast Cxt_1(s)) \oplus \ldots \oplus (Des_n(s,c) \circledast Cxt_n(s))$$

and then chooses a most desired control for execution. In practice, context dependent blending of behaviors is implemented by combining the output of all the behaviors using meta-rules of the form

$$\text{IF } Cxt_i \text{ THEN } apply(B_i)$$

and then defuzzifying with (1) to produce a tradeoff control (see Figure 1).

3 Control structures

Control structures, first introduced in [Ruspini, 1990], are the main ingredient for directing agent's activity. Each control structure acts as an elastic constraint: the fuzzy controller prefers the actions that best satisfy this constraint. Moreover, each control structure includes a specification of the conditions under which it is applicable. For example, a corridor may need to be followed only whenever the agent in inside it, and the path is clear.

More precisely, a control structure is a triple

$$S = \langle \mathrm{A}, \mathrm{B}, \mathrm{C} \rangle ,$$

where A is virtual object (an *artifact*) in the LPS; B is a behavior that specifies the way to react to the presence of this object; and C is the *context* where the control structure is relevant. An example of a control structure is

$$\langle \mathsf{CP1}, \mathsf{Go\text{-}To\text{-}CP}, \mathsf{near(CP1)} \rangle .$$

CP1 is a "control-point", a marker for a (x, y) location, together with a heading and a velocity; the associated behavior Go-To-CP reacts to the presence of a control point in the LPS by generating the commands to reach that location, heading and velocity; near(CP1) specifies that this behavior should be used only when the robot is sufficiently close to the control point. Hence, a control structure can be thought of as a specification of *what behavior* should be used with respect to *which object* and under *which conditions*.

Artifacts can correspond to real objects, like a wall to follow; these are normally placed in LPS based on prior information (e.g., map information), and are subsequently updated

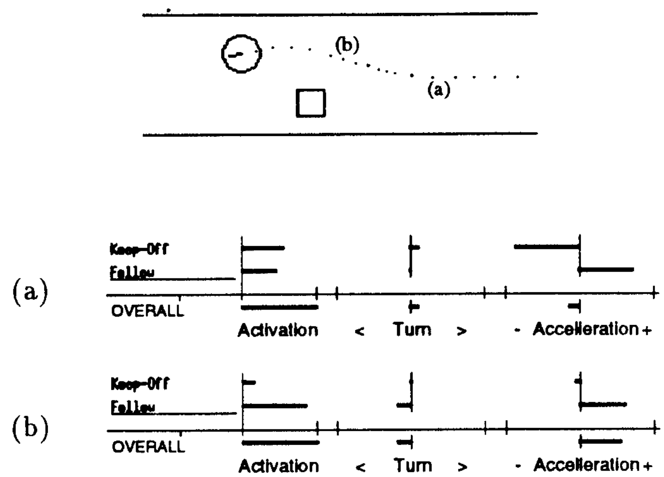


Figure 2: Blending reactivity with purposeful action.

on the basis of what is actually perceived. The combination of artifacts and fuzzy logic compensates for some uncertainty: the artifact provides an assumption for action when the sensors are not "seeing" the wall; the elasticity of fuzzy rules guarantees a smooth degradation of performance when this assumption is partially incorrect.

A set of control structures represents a complex control regime, or a *plan*. Depending from the current context, some of the control structures in the set will be more active, and some less. Context depending blending of the corresponding behaviors produces a overall control at each point. For instance, two consecutive *corridors to follow*, and a *door to cross*, together with the condition when each one is adequate, may constitute a plan for visiting an office. We have made experiments using traditional planning techniques to automatically generate sets of control structures for performing complex tasks [Saffiotti *et al.*, 1993a].

4 The outcome

The use of fuzzy logic in the definition of individual behaviors, as well as as a basis for behavior blending, has resulted in a robust mechanism for executing complex activities in the real world. In this section, we give some examples of the performance of this mechanism as implemented on our mobile robot, Flakey.

Flakey's controller includes behaviors, often called *reactive* in the robotic literature, whose goal is to promptly react to certain perceptual events in the LPS, like the KEEP-OFF behavior above. These behaviors are typically based on data that have undergone little or no interpretation, and hence very quickly available. Context-dependent blending of reactive

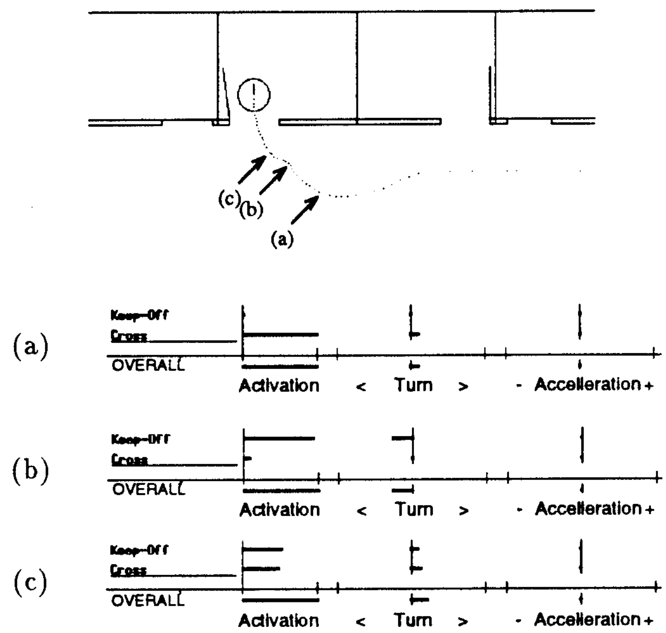


Figure 3: Compensating for inexact prior knowledge.

behaviors with purposeful ones provide Flakey with the ability to perform goal-oriented activities in an uncertain, dynamic environment. Figure 2 shows the result of blending the FOLLOW and the KEEP-OFF behaviors. The bars show the level of activation and the preferred controls (turn and acceleration) for each behavior, and the result of the blending.[4] In (a), an obstacle has been detected, and the preferences of KEEP-OFF are dominating; later, when the path is clear, the goal-oriented preferences expressed by FOLLOW re-gain importance (b).

Blending reactive and purposeful behaviors may also help in compensating for the imprecision of the prior knowledge. Figure 3 illustrates this point. In (a), the CROSS behavior is relying on prior information about the position of the door to cross. This estimate turns out to be off by some 40 centimeters, and Flakey is grossly misheaded. In (b), KEEP-OFF intervenes to avoid colliding with the edge of the door, re-orienting Flakey toward the door opening. Later (c), both behaviors cooperate to lead Flakey though the office door — i.e., through the perceived opening that is more or less in the assumed position.

Finally, Figure 4 shows an example of execution of a full plan. The plan consists of the

[4]The bar graph representation may be misleading: recall that each behavior actually generates a measure of desirability for *each* possible control, and that these measures —not just a preferred representative— are then used in the blending.

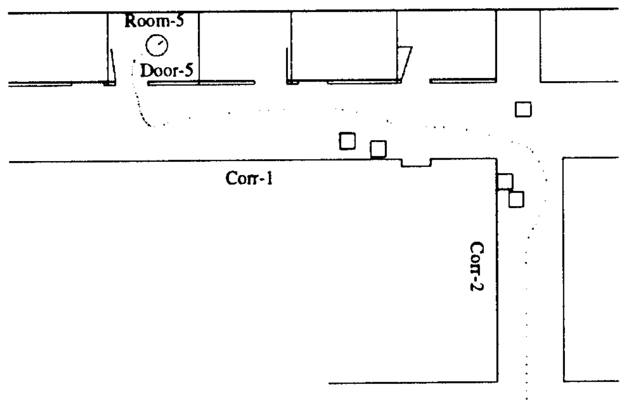


Figure 4: Robust execution of a full plan.

following four control structures[5]

$$\begin{aligned}
S1 &= \langle \text{Obstacle}, \textsc{Keep-Off}, \text{near(Obstacle)} \rangle \\
S2 &= \langle \text{Corr2}, \textsc{Follow}, \neg\text{near(Obstacle)} \wedge \text{at(Corr2)} \wedge \neg\text{near(Corr1)} \rangle \\
S3 &= \langle \text{Corr1}, \textsc{Follow}, \neg\text{near(Obstacle)} \wedge \text{at(Corr1)} \wedge \neg\text{near(Door5)} \rangle \\
S4 &= \langle \text{Door1}, \textsc{Cross}, \neg\text{near(Obstacle)} \wedge \text{near(Door5)} \rangle
\end{aligned}$$

This plan has been generated by a simple goal-regression planner based on a topological map annotated with approximate metric information. No obstacle was represented in the map. Total execution time was approximately 80 seconds, at top speeds of 400 mm/s.

5 Conclusions

We have proposed a behavior-based approach to autonomous execution of robot plans grounded in fuzzy logic. Our approach provides robustness in face of uncertain knowledge and unpredictable dynamics; principled combination of concurrent activities; and a simple, modular implementation. By ensuring reliable execution of high-level operation, our approach allows an agent to make effective use of coarse-grained plans generated by traditional AI techniques. From a more theoretical viewpoint, our study resulted in the development of the notion of *control structure* as a way to introduce high level symbolic goals into a fuzzy controller; and of context-dependent blending of behaviors as an effective technique for integrating multiple goals. The proposed technique has been successfully demonstrated on our mobile robot, Flakey.

[5]The actual plan has more control structures, including some for perceptual actions, and more complex contexts — see [Saffiotti *et al.*, 1993a].

Acknowledgments Nicolas Helft, John Lowrance, Daniela Musto and Leonard Wesley contributed to the development of the ideas presented in this paper. The first author has been supported by a grant from the National Council of Research of Italy. Research performed by the second author was supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620–91–C–0060. Support for Kurt Konolige came partially from ONR Contract No. N00014–89–C–0095. Additional support was provided by SRI International.

References

[Arkin, 1990] Arkin, R. C. 1990. The impact of cybernetics on the design of a mobile robot system: a case study. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6):1245–1257.

[Brooks, 1987] Brooks, R. 1987. Intelligence without representation. In *Procs. of the Workshop on the Foudations of AI*, MIT, Cambridge, MA.

[Congdon *et al.*, 1993] Congdon, C.; Huber, M.; Kortenkamp, D.; Konolige, K.; Myers, K.; Ruspini, E. H.; and Saffiotti, A. 1993. CARMEL vs. Flakey: A comparison of two winners. *AI Magazine* 14(1):49–57.

[Firby, 1987] Firby, J. R. 1987. An investigation into reactive planning in complex domains. In *Procs. of the AAAI Conf.*

[Gat, 1991] Gat, E. 1991. *Reliable Goal-Directed Reactive Control for Real-World Autonomous Mobile Robots.* Ph.D. Dissertation, Virginia Polytechnic Institute and State University.

[Hanks and Firby, 1990] Hanks, S. and Firby, R. J. 1990. Issues and architectures for planning and execution. In *Workshop on Innovative Approaches to Planning, Sheduling and Control*, San Diego, CA.

[Khatib, 1986] Khatib, O. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research* 5(1):90–98.

[Payton *et al.*, 1990] Payton, D. W.; Rosenblatt, J. K.; and Keirsey, D. M. 1990. Plan guided reaction. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6).

[Ruspini and Saffiotti, 1993] Ruspini, E. H. and Saffiotti, A. 1993. The formulation and solution of non-linear optimization problems in multi-valued logic. Technical report, SRI Artificial Intelligence Center, Menlo Park, California. Forthcoming.

[Ruspini, 1990] Ruspini, E. H. 1990. Fuzzy logic in the Flakey robot. In *Procs. of the Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA)*, Japan. 767–770.

[Ruspini, 1991a] Ruspini, E. H. 1991a. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning* 5:45–88.

[Ruspini, 1991b] Ruspini, E. H. 1991b. Truth as utility: A conceptual systesis. In *Procs. of the 7th Conf. on Uncertainty in Artificial Intelligence*, Los Angeles, CA.

[Saffiotti *et al.*, 1993a] Saffiotti, A.; Konolige, K.; and Ruspini, E. H. 1993a. Now, do it! Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti *et al.*, 1993b] Saffiotti, A.; RuspiniK., E. H.; and Konolige, 1993b. A fuzzy controller for flakey, an autonomous mobile robot. Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti, 1993] Saffiotti, A. 1993. Some notes on the integration of planning and reactivity in autonomous mobile robots. In *Procs. of the AAAI Spring Symposium on Foundations of Automatic Planning*, Stanford, CA. 122–126.

[Schoppers, 1987] Schoppers, M. J. 1987. Universal plans for reactive robots in unpredictable environments. In *Procs. of the Int. Joint Conf. on Artificial Intelligence.* 1039–1046.

[Suchman, 1987] Suchman, Lucy 1987. *Plans and situated actions: the problem of human machine communication.* Cambridge University Press, Cambridge, MA.

[Wilkins, 1988] Wilkins, D. E. 1988. *Practical Planning.* Morgan Kaufmann, San Mateo, CA.

A Fuzzy Controller for Flakey, an Autonomous Mobile Robot

Procs of the Fuzzy Logic 93 Conference
Burlingame. CA, 1993 (Penn Well Publisher)

A Fuzzy Controller for Flakey, an Autonomous Mobile Robot

Alessandro Saffiotti* **Enrique H. Ruspini** **Kurt Konolige**

Artificial Intelligence Center, SRI International
Menlo Park, CA 94025, U.S.A.
{saffiotti, ruspini, konolige}@ai.sri.com

Abstract

Controlling the movement of an autonomous mobile robot in real-world unstructured environments requires the ability to pursue strategic goals under conditions of uncertainty, incompleteness, and imprecision. We describe a fuzzy controller for a mobile robot that can take multiple goals into consideration. Through the use of fuzzy logic, goal-oriented behavior (e.g., trying to reach a given location) and reactive behavior (e.g., avoiding obstacles on the way) are smoothly blended into one sequence of control actions. The fuzzy controller has been implemented in the SRI robot Flakey.

1 Introduction

Many classical solutions to the problem of autonomous robot operation in artificial intelligence have been based on a "two level" paradigm, where a planner generates a sequences of operations whose performance is expected (in an ideal world!) to satisfy the robot's goals; and these operations are then executed by the robot. This approach has been criticized for not providing the real-time responsiveness, or *reactivity*, that is needed in real-world environments (for example, to avoid a person walking in front of the robot), and for being too inflexible in the face of uncertainty and imprecision in the information used (both prior knowledge, and perceptual information), and errors in the execution [Firby, 1987; Kaelbling, 1987; Gat, 1991; Saffiotti, 1993]. Some authors have proposed the use of fuzzy control techniques to overcome these limitations (e.g., [Sugeno and Nishida, 1985; Yen and Pfluger, 1992]). As they utilize continuous feedback, fuzzy controllers are inherently more reactive that the mainly open-loop control suggested in the two-level approach; and as they utilize fuzzy logic, they are inherently more tolerant to imprecision in knowledge and execution than crisp, purely symbolic approaches. However, fuzzy controllers are normally defined to achieve one specific, predetermined functionality of the system, or *goal*, even if this goal can be parameterized — e.g., to follow a path given at execution time. By contrast, an intelligent robot should be able to perform a variety of different tasks, and to consider

*On leave from Iridia, Université Libre de Bruxelles, Brussels, Belgium.

several goals simultaneously — e.g., to traverse a hallway while avoiding the obstacles and people on the way and while making sure that there is enough energy left.

We describe a fuzzy controller architecture that can achieve multiple symbolic goals, communicated by a planner or by a human programmer, and trade off conflicting goals. This architecture has been implemented in the SRI autonomous mobile robot, Flakey. Flakey's fuzzy controller implements robust high-level robot actions (like traversing a hallway, or crossing a door), and provides capabilities for:

1. Robust, uncertainty-tolerating goal-directed activity;
2. Real-time reactivity to unexpected contingencies; and
3. Blending of multiple goals.

The next section provides an outline of our mobile robot, and of the overall architecture of the fuzzy controller. Sections 3 and 4 describe how reactive and goal-oriented behaviors are produced, respectively. Section 5 discusses behavior blending and shows some examples. Finally, Section 6 concludes. An extended version of this note is available as [Saffiotti *et al.*, 1993b].

2 Architecture Overview

Flakey is a custom-built mobile robot platform approximately 1 meter high and .6 meter in diameter operating in an indoor environment. Two independently-driven wheels provide a maximum linear velocity of about .5 meters/sec. Sensors include a ring of 12 sonars, wheel encoders, and a video camera, currently used in combination with a laser to provide dense depth information over a small area in front of Flakey. A passive-vision system is currently being added. Flakey has enough computational power to run all the low-level and high-level interpretation and control processes on-board; in addition, Flakey's high-level processes can be run remotely through a radio link for better programming and debugging convenience.

Figure 1 sketches the architecture we developed for Flakey. The controller is centered on the notion of *behavior*. Intuitively, a behavior is one particular control regime that focuses on achieving one specific, predetermined goal, like avoiding obstacles, or reaching a specific location (given as a parameter). Hence, each behavior can be though of as being one separate parametric fuzzy controller. Behaviors take their input from a common storage structure, named *local perceptual space* (LPS), where all sensory input (possibly after some perceptual interpretation) is maintained. Moreover, the LPS may contain *control structures*, representing strategic goals to achieve, like a position to reach or a door to cross. Specific behaviors in the fuzzy controller react to the presence of control structures in the LPS by trying to satisfy them. Control structures can be put in the LPS by a human programmer, or by an automatic planner.

Many behaviors can be simultaneously active in the fuzzy controller. The outputs of all the currently active behaviors are combined together and a specific command is chosen for

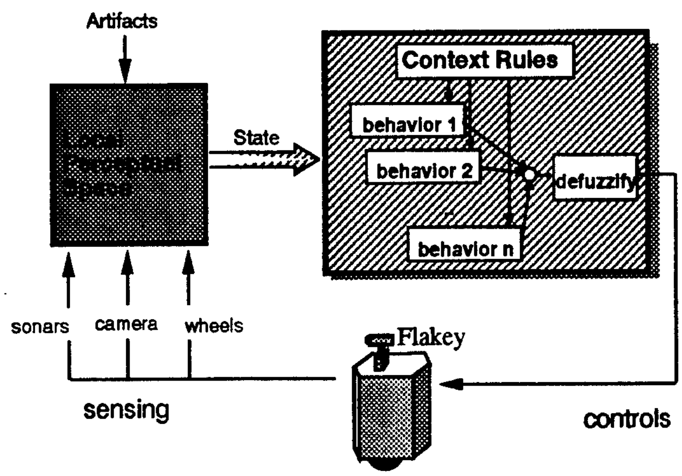


Figure 1: Flakey's architecture (partial).

execution ("defuzzification"). Before being combined, however, the output of each behavior is weighted according to how much that behavior applies to the current contextual situation — for example, a behavior for following a wall is scarcely applicable in the situation where there in an obstacle standing in front of Flakey, while one for avoiding obstacles is more adequate. The *context rules* encode the information about the applicability condition of each behavior.

The formal grounding of our behavioral approach, and of the context-depending blending of behaviors, rests on the notion of *desirability function* [Ruspini, 1990; Ruspini, 1991b]. Each behavior is represented by a function that maps each perceived situation to a measure of desirability of possible control from the point of view of that behavior's goal. Many behaviors, corresponding to many simultaneous goals, can be smoothly blended together by combining their desirability functions using the inferential procedures of fuzzy logic. The fuzzy controller prefers the actions that best satisfy each behavior. The formal perspective underlying our fuzzy controller is explored in detail in [Saffiotti *et al.*, 1993a].

3 Reactive Behaviors

Each *behavior* in the fuzzy controller is responsible for producing a certain type of movement, aimed at the attainment of certain goal. The simplest form of behaviors are the *reactive behaviors*: these map each input configuration, as in the LPS, to a control to apply in that situation. For example, a behavior for avoiding obstacles maps configurations of sonar readings where an obstacle is detected on the front left to the control of slowing down and turning right.

More precisely, and following [Ruspini, 1991b], we say that each behavior B is associated

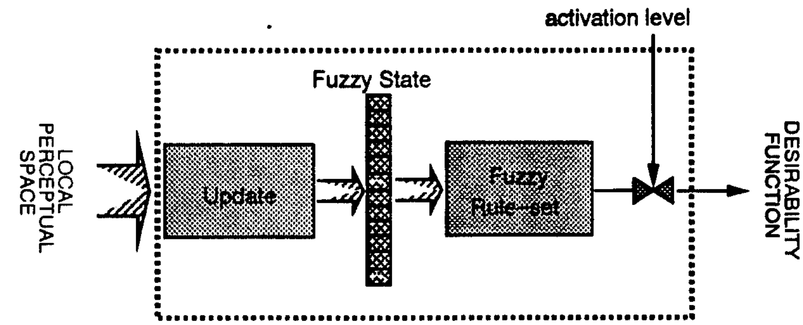


Figure 2: Implementation of a behavior.

with a desirability function

$$Des_B : \text{LPS} \times \text{Control} \rightarrow [0, 1]$$

that measures, for each configuration s of the LPS and value c of a control variable, the desirability $Des_B(s, c)$ of applying control values c in the situation s *from the point of view of* B. Equivalently, we can say that Des_B associates each situation s with the fuzzy set $\widetilde{C}$ of control values characterized by the membership function $\mu_{\widetilde{C}}(c) = Des_B(s, c)$. Notice that in general, c is a n-dimensional vector of values for all the control variables; in the case of Flakey, the control variables include linear acceleration and turning angle.

In practice, each behavior is implemented by a fuzzy machine structured as shown in Figure 2. The **Fuzzy State** is a vector of fuzzy variables (each having a value in $[0, 1]$) representing the truth values of a set of fuzzy propositions of interest (e.g., "obstacle-close-on-left"). At every cycle, the **Update** module looks at the (partially) interpreted perceptual input stored in LPS, and produces a new fuzzy state. The **Fuzzy Rule-Set** module contains a set of fuzzy rules of the form "If A then c" where A is a fuzzy expression composed by predicates in the fuzzy states plus the fuzzy connectives AND, OR and NOT; and c is a fuzzy set of values for the control variables. Max, min, and complement to 1 are used to compute the truth value of disjunction, conjunction and negation, respectively. An example of a control rule is:

```
IF obstacle-close-in-front
AND NOT obstacle-close-on-left
THEN turn sharp-left
```

Each "If A then c" rule computes the degree of desirability of applying the control values in c as a function of the degree at which the current state happens to be similar to A. The outputs of all the rules in a rule-set are unioned using the max T-conorm: the function computed in this way is meant to provide an approximation of the Des_B function above.[1]

[1]See [Ruspini, 1991a; Ruspini, 1991b] for an account of fuzzy logic and fuzzy control in terms of similarity and desirability measures, and the use of T-norms and T-conorms in this context.

This desirability function is fed to the **Defuzzify** module for computing one single control value. We presently do defuzzification according to the centroid approach: the resulting control value is given by

$$\frac{\int c\, Des_B(c)\, dc}{\int Des_B(c)\, dc}. \qquad (1)$$

4 GOAL-ORIENTED BEHAVIORS

The behaviors discussed in the previous section are purely reactive: at each cycle, Flakey selects an action solely on the basis of the current state of the world as perceived by its sensors and represented in the local perceptual space. Engaging into purposeful activities requires more than pure reactivity: we need to take explicit goals into consideration. For example, we may want Flakey to reach a given position at a given velocity, and still (reactively) avoid the obstacles on the way.

In our approach, a goal is represented by a *control structure*. Intuitively, a control structure is virtual object (an *artifact*) that we put in the LPS, associated with a behavior that encodes the way to react to the presence of this object. For example, a "control-point" is a marker for a (x, y) location, together with a heading and a velocity: the associated behavior GO-TO-CP reacts to the presence of a control point in the LPS by generating the commands to reach that position, heading and velocity. More precisely, a control structure is a triple

$$S = \langle A, D, C \rangle,$$

where A is an artifact, D is a desirability function that encodes the preferred relation between Flakey and the artifact, and C is a context of applicability (see below). Such a control structure implicitly defines a goal: the goal to achieve, and maintain, the relation D between Flakey and the artifact A.

Purposeful behaviors react to the presence of control structures in the LPS by generating a corresponding preference for controls. For example, a behavior for crossing a door reacts to the presence of a control structure

$$S1 = \langle \mathit{Wall1}, \textsc{Follow}, \mathit{clear\text{-}path} \rangle$$

in the LPS by generating preferences for the commands that keep Flakey parallel to the wall *Wall1* and at a fixed distance and proceeding at a given cruising speed. Hence, putting a control structure in the LPS is the basic way to communicate a goal to the fuzzy controller (provided that the controller includes a corresponding behavior).

Purposeful behaviors are implemented in the same form as reactive behaviors (see Figure 2 above), the only (formally invisible) difference being that the fuzzy state depends in general on properties of the artifact A of the control structure — e.g., its position relative to Flakey. For example, the FOLLOW-WALL behavior is meant to respond to the presence of the control structure $S1$ above; it includes the following rule:

```
IF too-close-on-right(wall)
THEN turn moderate-left
```

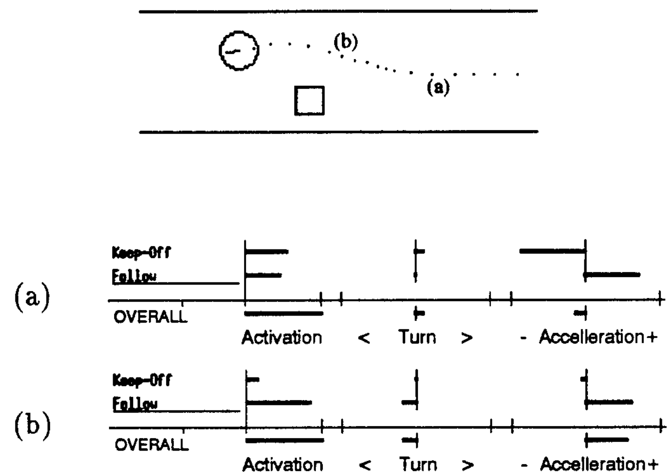


Figure 3: Blending reactive and purposeful behaviors.

where *wall* is the artifact of the control structure to which the behavior is reacting (in this case, *Wall1*).

5 Blending Behaviors

Many behaviors can be simultaneously active in the fuzzy controller, each aimed at one specific goal (see Figure 1). The fuzzy controller selects the controls that best satisfy all the active behaviors. However, not all behaviors are always applicable: for instance, the Follow-Wall behavior is most applicable when the wall is near and the path is clear; when there is an obstacle on the way, an obstacle avoidance behavior (called Keep-Off becomes more applicable. To account for this, we associate with each behavior B a *context of applicability*, expressed by a fuzzy predicate Cxt_B. If the behavior is reacting to a control structure, Cxt_B is simply the context C of the control structure (as is the case for the *clear-path* in the $S1$ control structure above); otherwise, it is defined as part of the behavior (as is the case for the Keep-Off behavior). Given n behaviors $\{B_1, \ldots, B_n\}$, the fuzzy controller combines their desirability functions, modulo their contexts, into one overall desirability function

$$Des(s,c) = (Des_1(s,c) \otimes Cxt_1(s)) \oplus \ldots \oplus (Des_n(s,c) \otimes Cxt_n(s))$$

and then chooses a most desired control for execution using the (1) above. We call this way of combining behaviors *context-dependent blending of behaviors.*

In practice, context dependent blending of behaviors is implemented by discounting the output of each behavior using context-rules of the form

$$\text{IF } Cxt_i \text{ THEN } apply(B_i)\,.$$

The truth value of the context Cxt_i is used as the **Activation level** input to the fuzzy

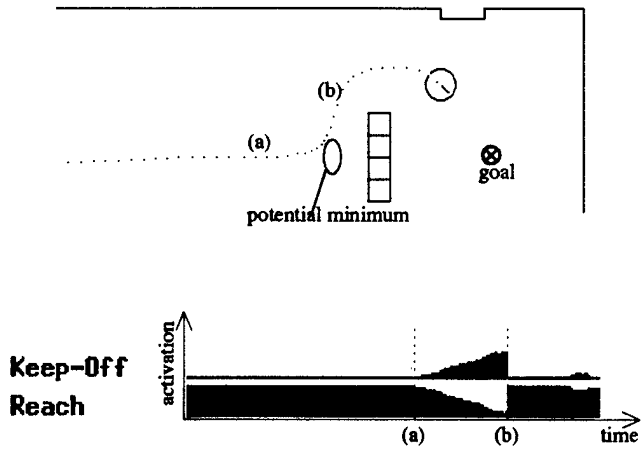


Figure 4: How context-dependent blending of behaviors helps avoiding local minima.

machine in Figure 2. The output of all the fuzzy machines, discounted by the corresponding activation levels, are then combined by $\oplus$ (MAX) and defuzzified by (1) to produce a tradeoff control (see Figure 1).

For example, consider the task of following a corridor while avoiding possible obstacles. The following context-rules express the desired interaction between the obstacle-avoidance behavior KEEP-OFF, and the corridor following behavior FOLLOW.

```
IF collision-danger
THEN APPLY(Keep-Off)

IF  NOT(collision-danger)
THEN APPLY(Follow)
```

Figure 3 shows the result of this blending. The bars show the activation level and the preferred controls (turn and acceleration) for each behavior, and the result of the blending. In (a), an obstacle has been detected, and the preferences of KEEP-OFF are dominating; later, when the path is clear, the goal-oriented preferences expressed by FOLLOW re-gain importance (b).

Figure 4 shows a similar example where a REACH behavior, meant to reach the goal point marked on the right, is blended with KEEP-OFF. The plot in the lower part shows the activation level of each beahvior over time. In (a), Flakey has perceived the obstacle; as the obstacle becomes nearer, the KEEP-OFF behavior becomes more active, at the expenses of the REACH behavior, and Flakey responds more and more to its suggestions (aimed at turning away from the obstacle), discarding the suggestions of the REACH behavior (aimed at heading toward the goal point). The REACH behavior then regains importance as soon as Flakey is out of danger (b), and Flakey resumes its goal-oriented course. Notice that

an un-weighted combination (e.g., the vector summation performed in most potential-field approaches to robot navigation [Khatib, 1986]) would result in the production of a zone of local equilibrium (local minimum) as indicated in the picture: when coming from the left edge, the robot would be first attracted and then trapped into this zone. By using contexts, the "attractive power" of the goal is gradually shaded away by the obstacle as Flakey approaches the obstacle, so Flakey responds more and more to the obstacle-avoidance suggestions alone.

6 Conclusions

The use of fuzzy logic in Flakey's controller has resulted in robustness in face of uncertain knowledge and unpredictable dynamics; principled combination of concurrent activities; and a simple, modular implementation. In return, our study resulted in the development of the notion of context-dependent blending of behaviors as an effective technique for integrating multiple goals in a controller; we have shown how this technique can be implemented in a two-level hierarchical rule-based system. A similar technique for dealing with multiple goals in a fuzzy controller has been previously proposed in [Berenji *et al.*, 1990]. Our solution extends this proposal by allowing the introduction of strategic goals in the controller; and by dynamically modify their degree of importance by the context mechanism.

Our controller presents several advantages over existing approaches to autonomous robot navigation. Firstly, the use of fuzzy logic at the movement control level results in improved robustness (e.g., more tolerance to sensor noise and knowledge imprecision), allowing our robot to make effective use of approximate and incomplete maps. Second, context dependent blending of behaviors provides a more principled approach to behavior combination when compared with other combination schemas (e.g., [Arkin, 1990; Payton *et al.*, 1990]). Finally, fuzzy rules appear to be a more powerful and natural way to express goals than the pseudo-potential functions required by the so-called "potential fields" methods, widely used for robot navigation (e.g., [Khatib, 1986]).

The performance of Flakey's controller has been demonstrated at the first AAAI robot competition in San Jose, CA [Congdon *et al.*, 1993]. The rules of the competition required that the robots perform purposeful activities in presence of unknown obstacles and moving people. Flakey accomplished the task and exhibited smooth movement and extremely reliable reactivity, as best summarized in one judge comment: "Only robot I felt I could sit or lie down in front of." (What he actually did!)

Acknowledgments John Lowrance, Daniela Musto, Karen Myers and Leonard Wesley contributed to the development of the ideas presented in this paper. Nicolas Helft implemented a first version of Flakey's controller. The first author has been supported by a grant from the National Council of Research of Italy. Research performed by the second author was supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620–91–C–0060. Support for Kurt Konolige came partially from ONR Contract No. N00014–89–C–0095. Additional support was provided by SRI International.

REFERENCES

[Arkin, 1990] Arkin, R. C. 1990. The impact of cybernetics on the design of a mobile robot system: a case study. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6):1245–1257.

[Berenji *et al.*, 1990] Berenji, H.; Chen, Y-Y.; Lee, C-C.; Jang, J-S.; and Murugesan, S. 1990. A hierarchical approach to designing approximate reasoning-based controllers for dynamic physical systems. In *Procs. of the 6th Conf. on Uncertainty in Artificial Intelligence*, Cambridge, MA.

[Congdon *et al.*, 1993] Congdon, C.; Huber, M.; Kortenkamp, D.; Bidlack, C.; Cohen, C.; Huffman, S.; Koss, F.; Raschke, U.; Weymuth, T.; Konolige, K.; Myers, K.; Saffiotti, A.; Ruspini, E. H.; and Musto, D. 1993. CARMEL vs. Flakey: A comparison of two winners. Technical report series, AAAI, Menlo Park, California. Short version appeared in *AI Magazine*, 14(1), Spring 1993, pp. 49–57.

[Firby, 1987] Firby, J. R. 1987. An investigation into reactive planning in complex domains. In *Procs. of the AAAI Conf.*

[Gat, 1991] Gat, E. 1991. *Reliable Goal-Directed Reactive Control for Real-World Autonomous Mobile Robots.* Ph.D. Dissertation, Virginia Polytechnic Institute and State University.

[Kaelbling, 1987] Kaelbling, L. P. 1987. An architecture for intelligent reactive systems. In Georgeff, M.P. and Lansky, A.L., editors 1987, *Reasoning about Actions and Plans.* Morgan Kaufmann.

[Khatib, 1986] Khatib, O. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research* 5(1):90–98.

[Payton *et al.*, 1990] Payton, D. W.; Rosenblatt, J. K.; and Keirsey, D. M. 1990. Plan guided reaction. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6).

[Ruspini, 1990] Ruspini, E. H. 1990. Fuzzy logic in the Flakey robot. In *Procs. of the Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA)*, Japan. 767–770.

[Ruspini, 1991a] Ruspini, E. H. 1991a. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning* 5:45–88.

[Ruspini, 1991b] Ruspini, E. H. 1991b. Truth as utility: A conceptual systesis. In *Procs. of the 7th Conf. on Uncertainty in Artificial Intelligence*, Los Angeles, CA.

[Saffiotti *et al.*, 1993a] Saffiotti, A.; Konolige, K.; and Ruspini, E. H. 1993a. Now, do it! Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti *et al.*, 1993b] Saffiotti, A.; RuspiniK., E. H.; and Konolige, 1993b. A fuzzy controller for flakey, an autonomous mobile robot. Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti, 1993] Saffiotti, A. 1993. Some notes on the integration of planning and reactivity in autonomous mobile robots. In *Procs. of the AAAI Spring Symposium on Foundations of Automatic Planning*, Stanford, CA. 122–126.

[Sugeno and Nishida, 1985] Sugeno, M. and Nishida, M. 1985. Fuzzy control of a model car. *Fuzzy Sets and Systems* 16:103–113.

[Yen and Pfluger, 1992] Yen, J. and Pfluger, N. 1992. A fuzzy logic based robot navigation system. In *Procs. of the AAAI Fall Symposium on Mobile Robot Navigation*, Boston, MA. 195–199.

Using Fuzzy Logic for Autonomous Vehicle Planning

Procs. of the NAFIPS Conference
Allentown, PA, August 1993

Using Fuzzy Logic for Autonomous Vehicle Motion Planning

Alessandro Saffiotti* Enrique H. Ruspini Kurt Konolige

Artificial Intelligence Center, SRI International
Menlo Park, CA 94025, U.S.A.
{saffiotti, ruspini, konolige}@ai.sri.com

Abstract

Autonomous vehicle operation in real-world unstructured environments requires capabilities for coping with uncertain, incomplete and approximate information in real time. We report on our progress on robust autonomous navigation using techniques based on fuzzy logic, and show some experiments performed on the SRI's mobile robot, Flakey.

1 Introduction

Autonomous operation of an unmanned vehicle (e.g., a mobile robot) in a real-world unstructured environment poses a series of problems. Firstly, the knowledge about the environment is in general incomplete, uncertain, and approximate. For example, maps typically omit many details and temporary features, things may have changed since the map was built, and the metric information available may be inexact. Secondly, perceptually acquired information is not reliable: sensor's noise introduces uncertainty and imprecision, sensor's limited range and visibility (e.g., occlusion) introduce incompleteness, and errors in the interpretation process may cause false beliefs. Thirdly, real-world environments have complex and largely unpredictable dynamics: objects can move, other agents may modify the environment, and environmental features may change (e.g., seasonal variations). Finally, vehicle's action execution is not completely reliable: the results produced by sending a given command to an effector can only be approximately estimated in general, and action execution may fail altogether.

In this paper, we report on our progress in the development of techniques for robust autonomous navigation of an experimental vehicle based on fuzzy logic. The techniques we describe have been tested on SRI's mobile robot, Flakey. However, their scope reaches beyond this particular testbed, as they attack such basic problems as the use of approximate and incomplete global maps to accommodate poor prior information; the integration of fuzzy control and symbolic reasoning (e.g., planning) to obtain robust goal-directed behavior; and

*On leave from Iridia, Université Libre de Bruxelles, Brussels, Belgium.

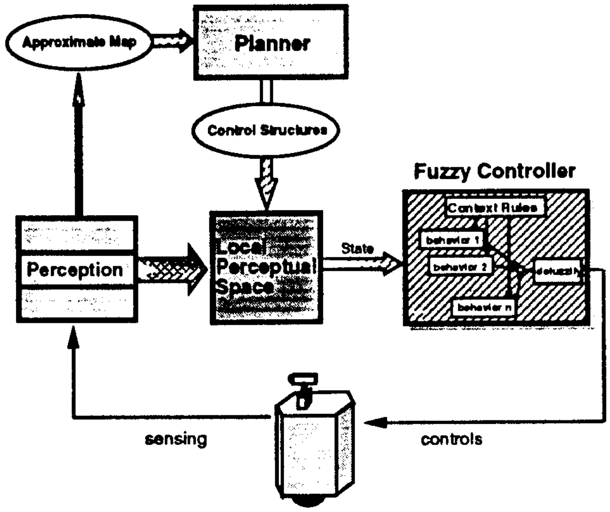


Figure 1: A simplified view of the architecture of Flakey.

the use of fuzzy logic for trading off possibly conflicting goals in real time — e.g., to achieve a given location while promptly reacting to unexpected events (like a person walking in front of the vehicle).

Figure 1 shows the architecture we developed for Flakey. The main data structure used during execution is the *Local Perceptual Space* (LPS), which provides a geometric picture of the space around the vehicle. LPS integrates information coming from the sensors, at different levels of abstraction (hence the different layers in the picture); and information coming from the planner. The latter consists of *control structures*, abstract descriptions of subgoals used by the fuzzy controller to orient the robot's activity. A set of control structures can be used to represent a decomposition of a complex task into more basic subtasks. Such a set, or *plan*, can be provided by a human programmer, or automatically generated by standard AI planning techniques [Saffiotti *et al.*, 1993a].

In the rest of this paper, we focus on two aspects of our work where the use of fuzzy logic is critical: movement control, and execution of complex navigation plans.

2 Movement Control

We base the low-level control of agent physical motion on a complex fuzzy controller (see Figure 1), whose role is to provide a layer of robust high-level motor skills. The controller is able to operate in real time under conditions of uncertainty, approximation and imprecision. The basic building block of this controller is a *behavior*. A behavior implements an atomic motor skill aimed at achieving or maintaining a given goal situation (e.g., to follow a given wall). Following Ruspini's semantic interpretation of fuzzy logic in terms of desirability

and utility [Ruspini, 1991a; Ruspini, 1991b], we represent each behavioral skill by means of a *desirability function* that expresses preferences over possible control actions from the perspective of that behavior's goal. (See [Saffiotti *et al.*, 1993b; Saffiotti *et al.*, 1993a] for a full account). For example, a behavior aimed at following a given wall prefers actions that keep the agent parallel to that wall and at a safe distance. Each behavior is implemented by a set of fuzzy rules of the form

$$\text{IF } A_i \text{ THEN } C_i$$

where A_i is composed of fuzzy predicates and fuzzy connectives,[1] and C_i is a fuzzy set of control vectors. For example, our robot platform, Flakey, includes a KEEP-OFF behavior, intended to keep the robot safely away from occupied areas (obstacles) as they are detected by the sonars. This behavior includes the following rule:

```
IF obstacle-close-in-front
AND NOT obstacle-close-on-left
THEN turn sharp-left
```

The controller can use any of the objects maintained in the LPS as its input. Purely reactive behaviors, intended to provide quick simple reactions to potential dangers (e.g., to avoid collisions) typically use sensor data that has undergone little or no interpretation. More purposeful behaviors, like reaching a given location, must take explicit goals into consideration. We represent goals into the LPS by means of *control structures*. A control structure is a triple

$$S = \langle \mathrm{A}, \mathrm{B}, \mathrm{C} \rangle \ ,$$

where A is virtual object (an *artifact*) in the LPS; B is a behavior that specifies the way to react to the presence of this object; and C is a fuzzy predicate expressing the *context* where the control structure is relevant. An example of a control structure is:

$$S1 = \langle \mathrm{CP1}, \text{Go-To-CP}, \mathit{near}(\mathrm{CP1}) \rangle .$$

CP1 is a "control-point", a marker for a (x, y) location, together with a heading and a velocity. The associated behavior GO-TO-CP reacts to the presence of S1 in the LPS by generating the commands to reach the location, heading and velocity specified by CP1. GO-TO-CP includes rules like:

```
IF  facing(CP1)
AND too-slow-for(CP1)
THEN accelerate smooth-positive
```

Finally, *near*(CP1) specifies that this behavior should be used only when the robot is sufficiently near to the control point — that is, these are the conditions under which the rules in the behaviors are expected to produce the intended result. Our fuzzy controller includes

[1] We use min, max, and complement to 1 for AND, OR, and NOT, respectively.

several purposeful behaviors — e.g., FOLLOW-WALL, FOLLOW-CORRIDOR, CROSS-DOOR, FACE-POINT, etc — each implemented by four to eight fuzzy rules.

Many behaviors can be simultaneously active in the fuzzy controller, each aimed at one specific goal. For instance, a purposeful behavior for following a wall can coexist with a reactive one for avoiding obstacles on the way. The fuzzy controller selects the controls that best satisfy all the active behaviors. Satisfaction is weighted by each behavior's relevance to the current situation, as measured by the truth value of the corresponding context predicate: for instance, the FOLLOW-WALL behavior is most applicable when the wall is near and the path is clear; while KEEP-OFF becomes more applicable when there is an obstacle on the way. Context dependent blending of behaviors is implemented by combining the output of all the behaviors using **Context-Rules** (cf. Figure 1) of the form

$$\text{IF } C_i \text{ THEN } apply(B_i).$$

The truth value of the context condition C_i in the current situation is used as an **activation level** to discount the output of the behavior B_i before combination (through the MAX operator). The result of the weighted combination of all the active behaviors is then defuzzified, and the command so obtained is sent to the vehicle for execution. This technique for combining multiple goals was originally inspired by the work of Berenji [Berenji *et al.*, 1990], and further developed by Ruspini [Ruspini, 1990].

3 Plan Execution

Context dependent blending of behaviors is our main mechanism for composing control structures into complex control regimes, or *plans*. Intelligent performance of complex and varied tasks by an autonomous agent requires that the agent itself be able to generate these plans. The use of fuzzy logic in our control structures allows us to level the disparity between the symbolic, discrete level of traditional planning and deliberating mechanisms developed in AI, and the analogical, continuous level of physical control. In this respect, control structures embody a declarative representation of executable actions: from the execution side, complex control regime can be expressed by set of control structures blended through the context mechanism exposed above. From the symbolic reasoning side, we have shown in [Saffiotti *et al.*, 1993a] how sets of control structures can be easily generated by traditional planning techniques.

Figure 2 shows an example of execution of a plan by Flakey. The core of the plan consists of the following four control structures:

$$\begin{aligned}
S1 &= \langle \text{Obstacle}, \text{KEEP-OFF}, \text{near(Obstacle)} \rangle \\
S2 &= \langle \text{Corr1}, \text{FOLLOW}, \neg\text{near(Obstacle)} \wedge \text{at(Corr1)} \wedge \neg\text{near(Corr2)} \rangle \\
S3 &= \langle \text{Corr2}, \text{FOLLOW}, \neg\text{near(Obstacle)} \wedge \text{at(Corr2)} \wedge \neg\text{near(Door5)} \rangle \\
S4 &= \langle \text{Door5}, \text{CROSS}, \neg\text{near(Obstacle)} \wedge \text{near(Door5)} \rangle
\end{aligned}$$

This plan has been generated by a simple goal-regression planner based on a topological map annotated with approximate metric information. No obstacle was represented in the map.

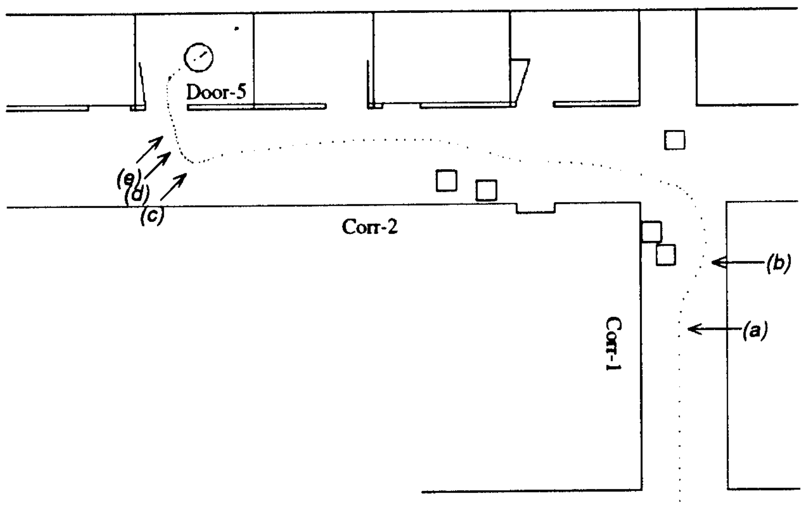


Figure 2: Robust execution of a plan.

Notice that plans are arbitrarily complex combinations of behaviors, possibly including information-gathering behaviors [Saffiotti, 1993]. This contrasts with other approaches to autonomous navigation based on fuzzy logic, where the agent generates and then follows a *path* (e.g., [Yen and Pfluger, 1992]).

At any point during the execution of the above plan, the KEEP-OFF control structure (S1) will become active should an obstacle be detected by the sensors. Blending S1 with the goal-oriented control structures S2–S4 results in Flakey pursuing its goals while smoothly going around obstacles (as in (a) and (b)). It also helps Flakey to compensate for imprecision in the map: in (c), $S4$ is relying on prior information to cross a door; as this turns out to be off by some 30 centimeters, KEEP-OFF intervenes to avoid colliding with the edge of the door (d); successively (e), both behaviors cooperate to lead Flakey though the opening which is more or less in the assumed position — e.g., through the actual doorway.

Figure 3 shows how behavior blending works in the case of FOLLOW and KEEP-OFF when avoiding the first obstacle in Figure 2 (a) and (b). The bars in the picture show, for each behavior, the level of activation and the preferred controls (turn and acceleration); the bottom level shows the result of the blending. In (a), the obstacle has been detected, and the preferences of KEEP-OFF dominate and cause Flakey to turn right and slow down; later, when the path is clear (b), the goal-oriented preferences expressed by FOLLOW regain importance and Flakey resumes its main course.

It is interesting to note that control structures also provides a mechanism to deal with the problem of self-localization in face of the approximate knowledge available in a map, ubiquitous in autonomous vehicle navigation in unstructured environments. Execution of

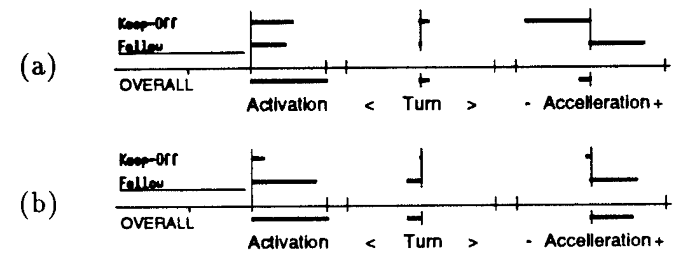


Figure 3: Blending reactive and purposeful behaviors.

control structures includes *anchoring* the artifact in the control structure to a physical object (see [Saffiotti *et al.*, 1993a] for more on this point). For instance, consider a FOLLOW action whose artifact is a given wall: the initial location of the artifact is extracted from the map. When the action is executed, perception is used to try to match a perceived wall-like object with this artifact. If the match is successful (i.e., the position and orientation are within certain tolerance ranges), the artifact is *anchored* to the perceived object, and its position is modified to reflect the position of the perceived object. The robot's motion is always performed relative to the artifact, which operates as an assumption when perceptual data is not available (e.g., when engaging in a new corridor, or the wall is occluded by a vacuum cleaner). The elasticity of fuzzy rules allows behaviors to tolerate errors in the assumed position, and guarantee a smooth degradation of performance when this assumption is incorrect.

The techniques discussed in this paper have been validated on many occasions. Flakey consistently performed innumerable runs inside the (unmodified) SRI's corridors and offices during normal working activity — the run shown above was one of these. Flakey's performance was also demonstrated at the first AAAI robotic competition (San Jose, CA, July 1992) [Congdon *et al.*, 1993]. The competing robots had to explore an unknown environment, recognize ten poles, and map their location. Later, they were required to return to three poles chosen by the judges. Flakey successfully completed the tasks, while avoiding obstacles and people, and placed overall second. Flakey gained special recognition for its smooth and reliable reactivity, as exemplified by one judge's comment: "Only robot I felt I could sit or lie down in front of." (He actually did!).

4 Conclusions

A solution to the problem of autonomous navigation in real world unstructured environment should include capabilities for both planned and real-time reactive behavior in situation of uncertainty and ignorance. The techniques that we have presented in this paper make use of fuzzy logic for providing flexible behavior that can tolerate imprecision in knowledge and execution, and for smoothly blending different behaviors aimed at different concurrent goals. We have implemented our techniques on an experimental indoor autonomous vehicle,

Flakey, and have integrated them with classical AI planning systems. Future work includes the automatic learning and improvement of fuzzy behaviors, the integration of more complex sensors (e.g., vision), and experiments in outdoor environments.

Acknowledgments Nicolas Helft, John Lowrance, Daniela Musto, Karen Myers, and Leonard Wesley contributed to the development of the ideas presented in this paper. A. Saffiotti was supported by a grant from the National Council of Research of Italy. E. Ruspini was supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620-91-C-0060. Support for K. Konolige came partially from ONR Contract No. N00014-89-C-0095. Additional support was provided by SRI International.

References

[Berenji *et al.*, 1990] Berenji, H.; Chen, Y-Y.; Lee, C-C.; Jang, J-S.; and Murugesan, S. 1990. A hierarchical approach to designing approximate reasoning-based controllers for dynamic physical systems. In *Procs. of the 6th Conf. on Uncertainty in Artificial Intelligence*, Cambridge, MA.

[Congdon *et al.*, 1993] Congdon, C.; Huber, M.; Kortenkamp, D.; Konolige, K.; Musto, D.; Myers, K.; Ruspini, E. H.; and Saffiotti, A. 1993. CARMEL vs. Flakey: A comparison of two winners. Technical report series, AAAI, Menlo Park, California. Short version appeared in *AI Magazine*, **14**(1), Spring 1993, pp. 49–57.

[Ruspini, 1990] Ruspini, E. H. 1990. Fuzzy logic in the Flakey robot. In *Procs. of the Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA)*, Japan. 767–770.

[Ruspini, 1991a] Ruspini, E. H. 1991a. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning* 5:45–88.

[Ruspini, 1991b] Ruspini, E. H. 1991b. Truth as utility: A conceptual systesis. In *Procs. of the 7th Conf. on Uncertainty in Artificial Intelligence*, Los Angeles, CA.

[Saffiotti *et al.*, 1993a] Saffiotti, A.; Konolige, K.; and Ruspini, E. H. 1993a. Now, do it! Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti *et al.*, 1993b] Saffiotti, A.; RuspiniK., E. H.; and Konolige, 1993b. A fuzzy controller for flakey, an autonomous mobile robot. Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti, 1993] Saffiotti, A. 1993. Some notes on the integration of planning and reactivity in autonomous mobile robots. In *Procs. of the AAAI Spring Symposium on Foundations of Automatic Planning*, Stanford, CA. 122–126.

[Yen and Pfluger, 1992] Yen, J. and Pfluger, N. 1992. A fuzzy logic based robot navigation system. In *Procs. of the AAAI Fall Symposium on Mobile Robot Navigation*, Boston, MA. 195–199.

Robust Control of a Mobile Robot using Fuzzy Logic

Procs of the First European Congress
on Fuzzy and Intelligent Technologies
EUFIT '93 (Aachen, Germany, Sept 1993)

Robust Control of a Mobile Robot using Fuzzy Logic

Alessandro Saffiotti * **Enrique H. Ruspini** **Kurt Konolige**

asaffio@ulb.ac.be ruspini@ai.sri.com konolige@ai.sri.com

IRIDIA
Université Libre de Bruxelles
50 av F. Roosevelt
B 1050 Bruxelles, Belgium

Artificial Intelligence Center
SRI International
333 Ravenswood ave
Menlo Park, CA 94025, U.S.A.

1 Introduction

We address the problem of controlling the motion of an autonomous mobile robot operating in a real-world unstructured environment. One of the major problems of this task is the extreme weakness of the diferent types of knowledge that need to be used. In the general case, the prior information available is approximate and incomplete; sensorial input is noisy and limited by the sensors' range and environmental features (e.g., occlusion); the dynamics of the environment is largely unpredictable (e.g., people moving around); and robot's actions are not completely reliable. Hence, a controller for an autonomous mobile robot must be able to operate under conditions of uncertainty, approximation and imprecision. Moreover, the robot must operate at the time-scale of the environment, and consider multiple goals simultaneously (e.g., reaching a position while avoiding obstacles and people). In order to cope with these difficulties, a controller must provide the robot with (at least) capabilities for

1. Real-time reactivity to unexpected contingencies;
2. Robust, uncertainty-tolerating goal-directed activity; and
3. Blending of multiple goals.

We have developed a controller based on fuzzy logic that satisfies these requirements, and tested it on the SRI mobile robot Flakey. In this note, we give an outline of this controller, and show how it satisfies the requirements stated above. A deeper description of Flakye's fuzzy controller can be found in [Saffiotti *et al.*, 1993b; Saffiotti *et al.*, 1993c].

*This work has been performed while the first author was visiting the AI Center of SRI International.

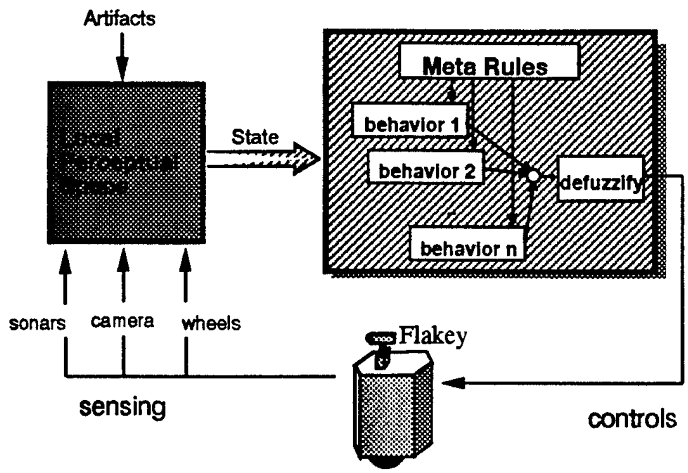


Figure 1: General architecture of the fuzzy controller.

2 The Fuzzy Controller

Figure 1 shows the architecture used. The controller is centered on the notion of *behavior*. Intuitively, a behavior is one particular control regime that focuses on achieving one specific, predetermined goal, like avoiding obstacles, or reaching a given location. Behaviors take their input from a common storage structure, named *local perceptual space* (LPS), where perceptual data, interpretations built from these data, and *artifacts* (discussed later) are maintained.

Following [Ruspini, 1990; Ruspini, 1991], we describe each behavior B in terms of a desirability function

$$Des_B : \text{LPS} \times \text{Control} \rightarrow [0,1]$$

that measures, for each state s of the LPS and control vector[1] c, the desirability $Des_B(s,c)$ of applying control c in the state s *from the point of view of* B. In every given state s, a behavior B induces an preferential order $\prec_B$ between possible controls in the obvious way: $c' \prec_B c''$ iff $Des_B(s,c') \leq Des_B(s,c'')$. For example, if the LPS state includes an obstacle on the left of Flakey, then right turning control actions will have higher desirability than left turning ones from the point of view of an obstacle avoidance behavior. The task of the controller is to compute $\prec_B$ at every step, and choose a corresponding maximally preferred control.

In practice, the controller is given approximations of desirability functions, written as sets of fuzzy rules of the form

$$\text{IF } A_i \text{ THEN } C_i \tag{1}$$

where A_i is a fuzzy formula composed of fuzzy predicates and fuzzy connectives,[2] and C_i is a fuzzy set of control vectors. Given a ruleset $\mathcal{R} = \{R_1, \ldots, R_n\}$, and a state s, the fuzzy controller first computes the desirability $Des_{\mathcal{R}}(s,c)$ of applying control c in state s according to the rules $\mathcal{R}$:

$$Des_{\mathcal{R}}(s,c) = (A_1(s) \otimes C_1(c)) \oplus \ldots \oplus (A_n(s) \otimes C_1(c)) \tag{2}$$

and then chooses one control $\hat{c}$ to apply, using centroid defuzzification:

$$\hat{c} = \frac{\int c \, Des_{\mathcal{R}}(s,c)\, dc}{\int Des_{\mathcal{R}}(s,c)\, dc}. \tag{3}$$

[1] In the case of Flakey, control vectors include linear acceleration and turning angle.

[2] In our implementation, we use min, max and complement to 1 for $\otimes$ (AND), $\oplus$ (OR), and $\ominus$ (NOT).

At each cycle of the controller, all the rules are evaluated in the current LPS state, and a new control vector is generated.[3] The perceptual processes that update the content of the LPS run asynchronously.

It is the task of the programmer to make sure that $Des_{\mathcal{R}}$ is a close enough approximation of the intended desirability function — i.e., that the fuzzy rules produce the expected behavior.

2.1 Reactivity

The controller includes behaviors whose goal is to react to certain perceptual events in the LPS. These behaviors are typically based on data that has undergone little or no interpretation, and hence very quickly available. For example, the KEEP-OFF behavior is intended to keep Flakey safely away from unknown obstacles as they are perceived by Flakey's sonars. A typical rule of KEEP-OFF is

```
IF   obstacle_close_in_front AND NOT obstacle_close_on_left
THEN turn sharp_left
```

The KEEP-OFF behavior includes four rules; the combination of these rules through (2) has been shown to produce effective maneuvers towards open areas, even in highly cluttered spaces.

2.2 Goal-directed activity

The controller includes "purposeful" behaviors that take explicit goals into consideration. Goals are represented by *artifacts* in LPS, as an imaginary line to follow: purposeful behaviors take these artifacts as input.[4] For instance, the FOLLOW-WALL behavior includes rules like

```
IF   wall_too_far_on_right
THEN turn moderate_right
```

Artifacts can correspond to real objects, like the wall above; these are normally placed in LPS based on prior information (e.g., map information), and are subsequently updated on the basis of what is actually perceived. The combination of artifacts and fuzzy logic compensates for some uncertainty: the artifact provides an assumption for action when the sensors are not "seeing" the wall; and the use of fuzzy logic guarantees a smooth degradation of performance when this assumption is incorrect.

2.3 Multiple goals

Many behaviors can be simultaneously active in the fuzzy controller, each aimed at one specific goal (see Figure 1): for instance, one for following a wall; one for keeping away from obstacles; and one for having the camera point at certain landmarks used for orientation. The fuzzy controller selects the controls that best satisfy all the active behaviors. However, not all behaviors are always applicable: for instance, the FOLLOW-WALL behavior is most applicable when the wall is near and the path is clear; while KEEP-OFF becomes more applicable when there is an obstacle on the way. Correspondingly, we associate to each behavior B a *context of applicability*, expressed by a fuzzy

[3] The cycle time of our controller is 100 msecs, which is adequate to our platform.

[4] Artifacts are typically provided by an external module, e.g., a planner.

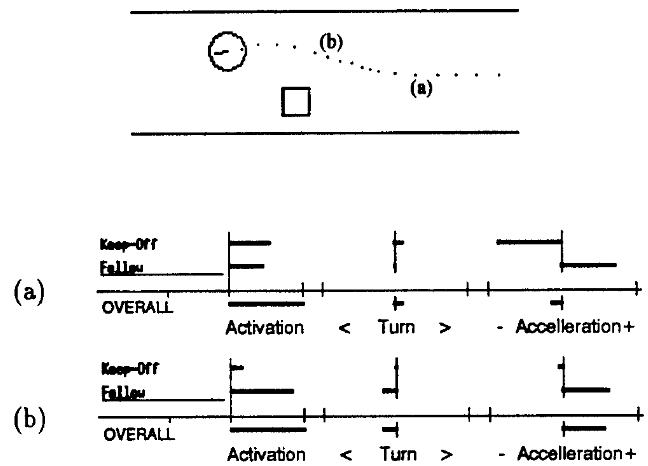


Figure 2: Blending reactive and purposeful behaviors.

predicate Cxt_B. Given n behaviors $\{B_1, \ldots, B_n\}$, the fuzzy controller combines their desirability functions, modulo their contexts, into one overall desirability function

$$Des(s,c) = (Des_1(s,c) \otimes Cxt_1(s)) \oplus \ldots \oplus (Des_n(s,c) \otimes Cxt_n(s)) \qquad (4)$$

and then chooses a most desired control for execution. We call (4) *context dependent blending of behaviors.*

In practice, context dependent blending of behaviors is implemented by weighting the output of each behaviors using meta-rules of the form

$$\text{IF } Cxt_i \text{ THEN } apply(B_i)$$

and then merging those outputs by a $\oplus$ T-norm (*max*, in our case), and defuzzifying with (3) to produce an overall tradeoff control (see Figure 1).

3 Examples

Context-dependent blending of behaviors is particular useful in combining purposeful action with the ability to react to unexpected contingencies. Figure 2 shows the result of blending the FOLLOW and the KEEP-OFF behaviors. The bars show the level of activation and the preferred controls (turn and acceleration) for each behavior, and the result of the blending. In (a), an obstacle has been detected, and the preferences of KEEP-OFF are dominating; later, when the path is clear, the goal-oriented preferences expressed by FOLLOW re-gain importance (b).

Blending reactive and purposeful behaviors can also help in compensating for imprecision in the prior knowledge. Figure 3 illustrates this point. In (a), the CROSS behavior is relying on prior information about the position of the door to cross. This estimate turns out to be off by some 40 centimeters (b), and KEEP-OFF intervenes to avoid colliding with the edge of the door. Later (c), both behaviors cooperate to lead Flakey though the actual doorway, by crossing the opening that is more or less in the assumed position.

Many purposeful behaviors can be blended to produce a complex activity. The context mechanism is used to decide when each behavior becomes relevant: plans including sequential and parallel activities can be built in this way. [Saffiotti, 1993; Saffiotti *et al.*, 1993a] duscuss the integration of Flakey's fuzzy controller with symbolic planning.

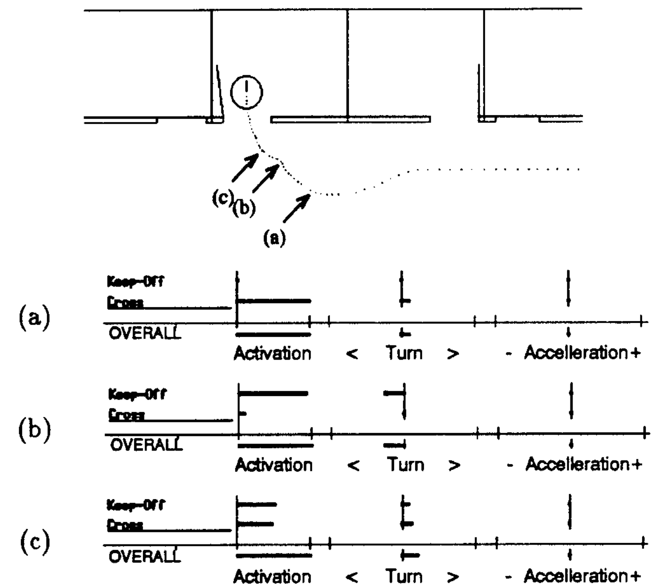


Figure 3: Compensating for inexact prior knowledge.

4 Conclusions

The use of fuzzy logic in Flakey's controller has resulted in robustness in face of uncertain knowledge and unpredictable dynamics; principled combination of concurrent activities; and a simple, modular implementation. In return, our study resulted in the development of the notion of context-dependent blending of behaviors as an effective technique for integrating multiple goals in a controller; we have shown how this technique can be implemented in a two-level hierarchical rule-based system. A similar technique for dealing with multiple goals in a fuzzy controller has been previously proposed in [Berenji *et al.*, 1990]. Our solution extends this proposal by allowing the introduction of strategic goals in the controller; and by dynamically modify their degree of importance by the context mechanism.

Our controller presents several advantages over existing approaches to autonomous robot navigation. Firstly, the use of fuzzy logic at the movement control level results in improved robustness (e.g., more tolerance to sensor noise and knowledge imprecision), allowing our robot to make effective use of approximate and incomplete maps. Second, context dependent blending of behaviors provides a more principled approach to behavior combination when compared with other combination schemas (e.g., [Arkin, 1990; Payton *et al.*, 1990]). Finally, fuzzy rules appear to be a more powerful and natural way to express partial goals than the energy function used in the so-called "potential fields" methods, used in many approaches to robot motion planning and control (e.g., [Khatib, 1986]).

The performance of Flakey's controller has been demonstrated at the first AAAI robot competition in San Jose, CA [Congdon *et al.*, 1993]. The rules of the competition required that the robots perform purposeful activities in presence of unknown obstacles and moving people. Flakey accomplished the task and exhibited smooth movement and extremely reliable reactivity, as best summarized in one judge comment: "Only robot I felt I could sit or lie down in front of." (What he actually did!)

Acknowledgments John Lowrance, Daniela Musto, Karen Myers and Leonard Wesley contributed to the development of the ideas presented in this paper. Nicolas Helft implemented an early version of Flakey's controller.

The first author has been supported by a grant from the National Council of Research of Italy. Research performed by Enrique Ruspini was supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620-91-C-0060. Support for Kurt Konolige came partially from ONR Contract No. N00014-89-C-0095. Additional support was provided by SRI International.

References

[Arkin, 1990] Arkin, Ronald C. 1990. The impact of cybernetics on the design of a mobile robot system: a case study. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6):1245-1257.

[Berenji *et al.*, 1990] Berenji, H.; Chen, Y-Y.; Lee, C-C.; Jang, J-S.; and Murugesan, S. 1990. A hierarchical approach to designing approximate reasoning-based controllers for dynamic physical systems. In *Procs. of the 6th Conf. on Uncertainty in Artificial Intelligence*, Cambridge, MA.

[Congdon *et al.*, 1993] Congdon, C.; Huber, M.; Kortenkamp, D.; Konolige, K.; Myers, K.; Ruspini, E. H.; and Saffiotti, A. 1993. CARMEL vs. Flakey: A comparison of two winners. *AI Magazine* 14(1):49-57.

[Khatib, 1986] Khatib, O. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research* 5(1):90-98.

[Payton *et al.*, 1990] Payton, D. W.; Rosenblatt, J. K.; and Keirsey, D. M. 1990. Plan guided reaction. *IEEE Trans. on Systems, Man, and Cybernetics* 20(6).

[Ruspini, 1990] Ruspini, E. H. 1990. Fuzzy logic in the Flakey robot. In *Procs. of the Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA)*, Japan. 767-770.

[Ruspini, 1991] Ruspini, E. H. 1991. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning* 5:45-88.

[Saffiotti *et al.*, 1993a] Saffiotti, A.; Konolige, K.; and Ruspini, E. H. 1993a. Now, do it! Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti *et al.*, 1993b] Saffiotti, A.; Ruspini, E. H.; and Konolige, K. 1993b. Integrating reactivity and goal-directedness in a fuzzy controller. In *Procs. of the 2nd Fuzzy-IEEE Conference*, San Francisco, CA.

[Saffiotti *et al.*, 1993c] Saffiotti, A.; RuspiniK., E. H.; and Konolige, 1993c. A fuzzy controller for flakey, an autonomous mobile robot. Technical report, SRI Artificial Intelligence Center, Menlo Park, California.

[Saffiotti, 1993] Saffiotti, A. 1993. Some notes on the integration of planning and reactivity in autonomous mobile robots. In *Procs. of the AAAI Spring Symposium on Foundations of Automatic Planning*, Stanford, CA. 122-126.